Professor Johan Paulsson **Dirk Landgraf**

Quantifying localizations and dynamics in single bacterial cells

Abstract

Levels of macromolecules fluctuate both spatially and temporally in individual cells. Such heterogeneity could be exploited for bet hedging in uncertain environments, or be suppressed by negative feedback if perturbations are deleterious. For the master stress-response regulator in *Escherichia coli*, RpoS, both of these scenarios have been suggested. RpoS levels are also exceedingly low and controlled by the ClpXP protease, which reportedly displays extreme spatial heterogeneity. However, little is known quantitatively about RpoS dynamics. This is partly because no functional protein fusions exist, but also because the quantitative tools for studying fluctuations and localizations are limited, particularly ones that can be independently validated. Here I develop such methods and begin applying them to RpoS.

Protein localization measurements increasingly rely on fluorescent protein fusions and are difficult to verify independently. I designed a non-intrusive method for validating localization patterns in live bacterial cells by exploiting post-division heterogeneity in downstream processes. Applying this assay to the ClpXP protease, widely reported to form biologically relevant foci, revealed in fact that the protease molecules are not specifically localized inside cells, as confirmed by four independent methods. I further evaluated 20+ commonly used fluorescent reporters and found that many cause severe mislocalization when fused to homo-oligomers, likely due to avidity effects. Further reinvestigating other foci-forming proteins strongly suggests that the previously reported foci were all caused by the fluorescent proteins used.

For mRNAs – which are often present in low numbers per cell and major sources of non-genetic heterogeneity – existing single-cell assays have unknown accuracy: the experimental counting errors could completely over-shadow the natural variation. I therefore optimized and cross-evaluated two single-molecule mRNA detection methods. Several problems were identified and solutions discussed.

I succeeded in building a functional RpoS protein fusion, and used bulk methods to show that the RpoS feedback loop is effectively not operating during exponential-phase growth. Mathematical analyses and initial experiments in a microfluidic device further suggest that the RpoS system has several unusual properties contributing towards extremely fast stress response. A stochastic analysis further suggests that the RpoS feedback loop cannot suppress spontaneous fluctuations, and preliminary experiments indicate that large deviations might indeed play important roles.

Table of Contents

Acknowledgement

Many years ago, after I finished high school in Germany, my dream at the time was to do a PhD in Biology at MIT. Little did I know at that time, but reality would be even better. I am especially grateful to the Harvard Systems Biology Department and the PhD program to have given me this tremendous opportunity. It was exciting to work in such a remarkable and interdisciplinary environment, and to be able to interact with so many talented and gifted people, who are all excited and enthusiastic about quantitative biology.

I would particularly like to thank the members of the Paulsson lab with whom I interacted on a daily basis. It was a fun and very interdisciplinary environment, where people helped each other and learned from each other. I also enjoyed witnessing how physicists and alike learned “swimming” in a wet lab and how random pieces of DNA “magically” matured into functional gene fusions.
It was a great pleasure to have worked with Burak Okumus, who introduced me to single-molecule imaging, taught me many things about microscopy and general organization. I usually paid back his support by “lending” him every two years the “Klose” jersey for a few days.
Also, I would like to thank Andreas Hilfinger for answering many math-related questions, engaging in encouraging scientific discussions and for just being Andy.
Further, I would like to thank Dann Huh for many helpful discussions, sharing ideas, code, reagents, and finding the answer to the question: “Where is Andy?”
I would like to thank Billy Lau for being a great cubical and lab mate and teaching me the real meaning of the word “easily”, the secrets about “Canadian laws”, and making my bench always look clean & organized.
I am thankful to Meriem El Karoui for help with paper writing and many fruitful scientific discussions.
I would like to thank Nate Lord for help with the mother machine, providing plasmids and sharing project ideas.
I am grateful to Shay Tal for many scientific discussions and discourses and also for not being upset with me despite the fact that I use (i) Nikon microscopes, (ii) Apple computers, (iii) believe in FISH and (iv) missed over a thousand times the communal 12 o’clock lunch.
I would like to thank Per Malkus for doing all general lab organization, ordering, and keeping things running on a daily basis.
Further I would like to thank all the other present and past members of the Paulsson lab: Raul Fernandez-Lopez, Rishi Jajoo, Tom Norman, Juan Pedraza, Yoonseok Jung, Ghee Chuan Lai, Anna Andersson and Jeanne Salje for being great co-workers.

During my PhD, I benefited tremendously from having an outstanding dissertation advisory committee. Every committee meeting contributed to progression of my projects and provided me with a big motivational boost. I am especially thankful to Tim Mitchison for sharing his insight, intuition, wisdom and spreading his excitement about Biology. It was a tremendous pleasure to have Alexander van Oudenaarden be part of my committee. I also would like to thank Alexander for first introducing me to Systems biology (in the fall of 2005 when I audited his course at MIT). Further, I would like to thank Erin O’Shea for her great support and many excellent suggestions during the meetings.

During my PhD, I was involved in many collaborations, inside and outside our laboratory. The following people contributed to the work that is presented in chapter 2. Burak Okumus built the single-molecule imaging setup and helped me with the HILO imaging experiments. Raul Fernandez-Lopez contributed the data about SeqA-mGFPmut3. Shay Tal provided the plasmids for the *lacO* array experiments. Peter Chien and Tania Baker contributed antibodies against ClpX and ClpP and gave many valuable comments and suggestions. I would also like to acknowledge Dann Huh for providing the mGFPmut3 variant (the first GFP that did not form foci when fused to ClpP).
Maier Avendano contributed greatly to the work presented in chapter 3. Maier helped me with the *in vitro* FISH assays and PP7 experiments, and he performed the *clpX* mRNA *in vitro* transcription. I would also like to thank Ralf Jungmann for help with our initial super-resolution imaging experiments. I am also grateful to Robert Singer for providing FISH probes and Daniel Zenklusen for protocols.
The work presented in chapter 4 started as a collaboration with Celeste Peterson and Tom Silhavy and they provided strains and antibodies. I would also like to acknowledge the help of Nate Lord with the mother machine experiments. He designed and built the chips and taught me some microfluidics.

I would further like to thank Pam Silver for providing me with the opportunity to carry out the research of my Master's thesis in her lab (in 2005) and for not leaving me behind at Dana-Farber when she moved to the Systems Biology Department.
I am grateful to David Rudner for many interesting discussions about the Clp foci and other scientific topics.
I would also like to thank Xindan Wang for help with *B. subtilis* experiments
I would like to thank Jennifer Waters for first introducing me to microscopy and providing constant technical support regarding imaging.
I would like to thank Becky Ward for giving us comments on our paper and also for writing a blog entry about my work.
I would also like to thank Samantha Reed for performing many administrative tasks, helping with visa applications and scheduling all committee meetings.
Further, I would like to thank the following people from the department: Martin Loose, Ian Swinburne, Tobias Bollenbach, and Erdal Toprak for help with techniques.

I would also like to thank my fiancée Oona Johnstone for constant help and encouragement during my PhD.
I am also very grateful to my parents for their support during my PhD and before I started graduate school, and for always encouraging me to do what I am interested in.

Finally, I would like to thank my PhD advisor Johan Paulsson for his constant scientific support and tremendous help over the last 5–6 years. It was a great pleasure working and interacting with him. His scientific approach to biology is unique and very interdisciplinary. It was a true Systems Biology experience and I learned a great deal.
I am very thankful that he gave me the freedom to work independently and follow my own ideas. Johan created a very friendly, open, and democratic lab environment and it was a great pleasure being a part of it.

I. General introduction

Many mRNAs and proteins are present in low numbers per cell, randomly collide and react, and segregate stochastically at cell division. Fluctuations in one component can also spread through interaction networks and affect downstream components, potentially creating non-genetic variability in any cellular process.

Intracellular concentrations can also fluctuate in space due to inhomogeneities. Even without organelles or other compartments, macromolecules can form large complexes that localize in cells and hence create local concentrations. Spatial heterogeneity can then create cell-to-cell variability when cells divide because of uneven portioning of the localized molecules. For example, for molecules that localize in a single focus per cell, segregation of this focus to one daughter but not the other creates large differences between the cells after cell division. This could have major effects on cellular physiology.

Single cells are in some sense unique individuals and this individuality cannot be captured by bulk measurements that average over many cells. Analyses of single cells have indeed revealed numerous cellular behaviors[1-6] that are impossible to study with bulk measurements. Whether explicitly focusing on cellular heterogeneity or studying cellular control systems, single cell experiments are key.

I believe that most fundamental questions are, at least initially, best studied in the simplest model system that still exhibits the properties of interest. In many ways, the bacterium *Escherichia coli* (*E. coli*) is therefore a natural starting point for addressing basic questions of cell-to-cell heterogeneity. *E. coli* is arguably the most thoroughly characterized model organism, perhaps alongside *Saccharomyces cerevisiae. E. coli* cells are unicellular and easy to culture and genetically modify, and their low

autofluorescence is advantageous in quantitative imaging experiments, allowing for single-molecule detection.

A system that deserves to be a model system for fluctuations, though little work has been done in single cells, is the regulatory network controlling bacterial stress response. Many of the key proteins were recently shown to be present in exceedingly low numbers per cell and fluctuations may then be used for bet hedging.

The master regulator of the bacterial stress response and stationary phase is the alternative sigma factor RpoS, also called σ^S, which controls the expression of over 500 genes in *E. coli*[7]. RpoS is extremely short-lived and present in very low numbers in non-stressed cells. A complicated network tightly regulates the number of RpoS molecules per cell at the level of transcription, translation and protein degradation[8,9]. There is experimental evidence that RpoS is also self-controlled by a negative feedback loop via regulated degradation[10,11], wherein RpoS activates the transcription of *sprE* which encodes an adapter protein, and SprE delivers RpoS to the ClpXP protease for degradation. This type of negative feedback loop is a very common motif occurring in diverse systems such as p53-Mdm2[3,12], NFκB-IκB[13] and the heat shock response[14,15].

It has been suggested in the literature[8,10,11,16] that the role of the RpoS-SprE feedback loop is to prevent haphazard and detrimental expression of RpoS, largely based on the fact that negative feedback loops are generally associated with maintaining homeostasis and correcting fluctuations. From first principles, however, heterogeneity in stress response could also be advantageous. In fact, having a few cells in a population that respond quickly to a stress could act as an insurance policy if conditions deteriorate so rapidly that the bulk of the population did not have a chance to seek protection in time.

Considering only qualitative aspects of the RpoS-SprE negative feedback loop, it seems plausible that its role is to maintain homeostasis and decrease fluctuations in RpoS. However, negative feedback can either suppress or amplify fluctuations depending on the exact parameters and molecular mechanisms[17]. To help to suppress noise, signaling molecules (SprE) must be made at high rates and have short lifetimes relative to the component that is controlled (i.e. RpoS). However, parameters reported in the literature and experimentally measured by our collaborators (C. Peterson and T. Silhavy) suggest that this is not the case here: SprE is made at extremely low rates and has a much longer half-life than RpoS. Noise suppression is then virtually impossible, unless fluctuations come from slow changes in environmental inputs to RpoS. These observations motivated the mathematical and experimental research presented here, which is focused on determining what role the RpoS-SprE negative feedback loop plays under normal conditions and in stress response.

When a cell is exposed to stress it is important to respond quickly. Because negative feedback loops in principle can speed up the adjustment to new steady states, the RpoS-SprE feedback loop could thus help orchestrate a faster target gene response than without feedback. This hypothesis was also investigated here, both mathematically and experimentally.

Despite the great importance of RpoS for cell survival and resistance to external stresses[18], and the key roles fluctuations are likely to play, there is virtually no quantitative study about RpoS dynamics, in bulk or in single cells. Perhaps the main reason is that functional protein fusions to RpoS are non-existent. To our knowledge, all attempts to add reporter tags to RpoS completely abolished RpoS activity. I therefore constructed and evaluated various fluorescent protein fusions to RpoS and SprE, and obtained functional protein fusions that now allow visualization of RpoS dynamics in live cells (see **Chapter 4**).

During exponential growth, RpoS is rapidly degraded by the ClpXP protease, which ensures an RpoS half-life of ~1–2 minutes and very low RpoS levels. In the last few years, several studies have reported that the Clp proteases are spatially localized, typically forming a single bright focus per cell but some cells have a dim focus or even no focus at all. This could have a major effect on cell-to-cell variability in RpoS levels, since some cells would degrade RpoS quickly and other cells would barely degrade RpoS at all. Cells with no protease foci could even temporally stabilize RpoS to the point where RpoS turns on stress response genes. Thus, the protease foci are a natural candidate for a bet-hedging mechanism. Bacteria that use a bet-hedging strategy produce populations with slightly different phenotypes to maximize survival and growth under changing conditions[19].

To address the question of whether the Clp protease foci generate cell-to-cell variability, I further characterized the central proteolysis machinery ClpXP in *E. coli*, with quantitative imaging techniques. This analysis is relevant far beyond RpoS. Numerous studies across at least five bacterial species (*E. coli*[20-22], *Bacillus subtilis*[23-27], *Caulobacter crescentus*[28,29], *Mycobacterium tuberculosis*[30] and *Mycobacterium smegmatis*[30]) have reported that these protease complexes form localized foci, and the foci were argued to have important roles for a wide range of biological processes. The Clp proteases have become a model system to demonstrate how bacteria localize central machineries, though virtually no studies have characterized the dynamic properties of the proteolysis system in single cells.

I therefore developed an experimental platform to characterize proteolysis in single *E. coli* cells and to analyze the variability in degradation rates after cell division. Exploiting these fluctuations as a single-cell analysis tool surprisingly revealed that the previously reported foci do not reflect the native localization of the Clp proteases. Protein

localization measurements are usually very difficult to validate because only two methods (i.e. immunofluorescence and fluorescent protein fusions) exist and both have rather strong limitations. The fluctuation-based test in combination with other experiments showed that the Clp protease foci are in fact artifacts caused by the fluorescent protein tags (Landgraf *et al.*, Nature Methods, 2012). I verified this finding with four independent methods. I then analyzed over twenty fluorescent reporters and found that many commonly used fluorescent proteins can cause severe mislocalization when fused to homo-oligomers or molecules that cluster. Multimerization of fluorescent proteins can be greatly reduced by the A206K mutation[31] and introducing this mutation into superfolder GFP (sfGFP), currently the brightest GFP variant on the market, converted sfGFP from a strongly foci-inducing tag to a monomeric reporter (msfGFP). Together with other members of the Paulsson lab, I also found that foci formation of many other proteins were either caused or greatly aggravated by sticky fluorescent protein tags (see **Chapter 2**). This ruled out the possibility that large RpoS fluctuations are caused by a localized ClpXP proteolysis system.

Since all RpoS fusion proteins reported so far have abolished RpoS activity, an alternative approach to experimentally analyze fluctuations would be to monitor fluctuations in the mRNA levels of *rpoS*, *sprE* and downstream RpoS-regulated target genes. This would only provide an imperfect proxy for the relevant protein fluctuations, but can still reveal much information about the network. Just as for the localization study, however, this requires mRNA counting methods that can be independently verified. The challenge when studying single-cell heterogeneity is that many results are difficult to interpret because the observed distributions are unknown mixes of real biological heterogeneity and measurement errors. Reliable quantitative reporters and assays are

therefore of great importance. For example, mRNA molecules can be detected in single cells with various methods that have single-molecule sensitivity, but most such methods seem to have far from single-molecule accuracy. The only study that reported controls suggesting that the counting results are quantitative[32] had to first modify the transcripts to the point where they were no longer degraded and barely moved in cells. The problem is that in smaller cells and at higher abundances, the point-spread function of standard light microscopy prevents close targets (< 200 nm) from being spatially separated, and it is instead necessary to quantify the total fluorescence intensity. If the assays were quantitative, the histogram of total intensities should then display equidistant peaks corresponding to the integer number of molecules. However, even in the most careful and accurate studies I have seen[33], the reported histogram did not display any peaks. In many studies it is therefore unknown if the results from mRNA counting methods reflect the natural heterogeneity or if the observed distributions are dominated by measurement noise.

After reviewing currently available methods, I report the optimization and evaluation of two independent methods, based on single-molecule mRNA FISH and RNA-binding protein reporters for live-cell mRNA imaging in bacteria. The two approaches are set up to allow for cross-validation of the assays to facilitate trouble-shooting. The analyses identified steps where conventional procedures could greatly reduce the accuracy of the assay, but also suggest potentially straightforward solutions (see **Chapter 3**). I am hopeful that completing this work will produce two working assays that can be independently verified.

In summary, this thesis reports substantial progress towards analyzing the master regulator of bacterial stress response in *E. coli*. Perhaps more importantly, by requiring that all key methods are independently verified – not by merely adjusting the parameters

of one protocol to match the results of the other, but by more rigorously demonstrating that the assays are quantitative – the work reveals surprising and important features of protein localization in bacteria, and to some extent about the reliability of mRNA counting methods. I believe these methods set the stage for more quantitative analyses of the bacterial stress response system, and hope they will be used more broadly.

II. Segregation of molecules at cell division reveals native protein localization

2.1. Introduction

2.1.1. Despite their small size and perceived simplicity, bacterial cells exhibit spatial organization

An average *Escherichia coli* (*E. coli*) cell, growing in rich medium, has a length of ~3.5 μm and a width of ~1.4 μm[34]. The volume is thus roughly a hundred to thousand times smaller than a typical mammalian cell[35] and also less spatially structured, without a nucleus or membrane-bound organelles. The cytoplasm is separated from the environment by an inner and an outer membrane as well as a rigid cell wall, and contains copies of the bacterial chromosome densely packed into irregular-shaped nucleoids.

Despite this apparent simplicity, a few well-studied examples[36,37] revealed that bacteria, like eukaryotic cells, are spatially organized and localize proteins to specific intra-cellular sites. For example, the prokaryotic tubulin homolog FtsZ localizes to the cell division septum of a dividing *E. coli* cell, as shown in 1991 with immunogold labeling and electron microscopy[38]. This finding has since been verified and refined by many groups, using various techniques including epi-fluorescence microscopy of a FtsZ fusion to green fluorescent protein (GFP)[39] or of FtsZ-tagged with a single fluorescent amino acid[40] and super-resolution imaging with photo-convertible fluorescent protein fusions[41,42]. The contractile FtsZ ring has even been reconstituted *in vitro* in liposomes with membrane-tethered FtsZ[43]. Other well-studied examples of spatially organized

proteins include the Min system that exhibits a spatial oscillation between the poles of the cell to determine the center point[44], and SeqA that binds in large numbers to the *E. coli* chromosome[45].

2.1.2. Existing methods to study protein localization in cells

The available methods to study protein localization in cells are practically limited to immunofluorescence (IF) or fusions to fluorescent proteins (or small fluorescent tags[46]). Immunofluorescence[47] was historically the primary method[48,49] but it is being phased out owing to its labor-intense procedures, the requirement for very specific antibodies, and the potential for fixation artifacts such as sensitivity to protocol parameters that affect the accessibility of the epitope or penetration of the antibody[50]. Immunofluorescence is further limited to dead cells and hence cannot be used to study protein dynamics in real time. The main advantage of immunofluorescence is that detection of the protein of interest requires no tag or modification. It thus allows for visualization of the native proteins, but in the non-native context of a chemically altered cell.

Fluorescent proteins (FPs), like GFP (**Figure 2.1**), offer many advantages over immunofluorescence and have quickly became the method of choice to study protein localization. This is partly because FP fusions work in live cells[51,52] and thus allow the study of dynamics[53,54], partly because no chemical treatment of the cells – which could introduce artifacts – is necessary, and partly because FPs are so convenient and easy to use. For these reasons FPs have been used on a genome-scale to analyze sub-cellular protein localization in many model systems like *E. coli*[20,21], *Caulobacter crescentus* (*C. crescentus*)[28], *Schizosaccharomyces pombe* (*S. pombe*)[55-57], *Saccharomyces cerevisiae* (*S. cerevisiae*)[58], *Arabidopsis thaliana*[59], *Nicotiana benthamiana*[60], *Drosophila*[61], body wall muscle cells of *Caenorhabditis elegans*[62], kidney cells of the African green monkey[63], mouse fibroblast cells[64] and human cells[65].

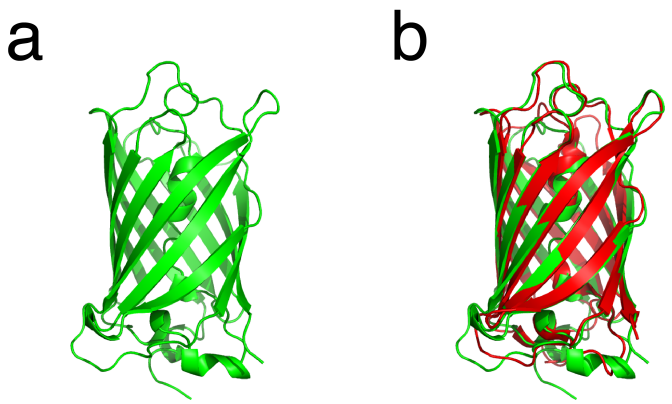


Figure 2.1. Green fluorescent protein (GFP) is a 238 amino acid long protein with a molecular weight of ~27 kDa. (**a**) The three-dimensional structure of GFP was solved by x-ray crystallography[66,67] in 1996 and is reminiscent of a cylinder with the chromophore buried in the inside. The GFP isolated from jellyfish *Aequorea victoria* is often referred to as "wild-type" GFP. The structure displayed here corresponds to protein data bank (PDB) entry 1GFL. (**b**) Isolation of fluorescent proteins from other sources, like Anthozoa (sea anemones and corals), and excessive protein engineering[68,69] resulted in various color-variants, like the red fluorescent protein mCherry[70]. All characterized fluorescent proteins share the same structural motif, despite very low sequence similarity and different origins. Overlay of the GFP and mCherry three-dimensional structures (GFP in green, mCherry in red). The PDB number of the mCherry structure is 2H5Q.

2.1.3. Many bacterial proteins are reported to localize in cells

The three main bacterial fluorescent protein fusion libraries[20,21,28], using three different types of fluorescent proteins, report that 8–15% of proteins significantly localize in foci (**Table 2.1**), and many also display other types of localization patterns such as septum or membrane localization. This has contributed to the notion that bacteria are far from 'bags of enzymes' and that spatial localization plays important roles in a wide range of processes. However, the addition of the fluorescent protein tag to the protein of interest can invoke potential problems, such as for example, changing the function[71], expression level[72], half-life[73,74], incorporation into a macromolecular assembly[74], the localization[58] or even induce artifactual helix formation[75] of the fusion compared to the untagged

endogenous protein. Appropriate controls[76] are therefore necessary but only rarely reported.

Table 2.1

Library name	Fluorescent protein	Organism	Comments	Foci
ASKA (A Complete Set of *E. coli* K-12 ORF Archive) library [20]	GFPuv4	*E. coli*	Fusion expressed from plasmid, inducible promoter, endogenous gene untagged	663 out of 4351 proteins (15.2%)
Yellow fluorescent protein fusion library for *E. coli* [21]	Venus YFP	*E. coli*	Chromosomal integration at endogenous gene locus, native promoter, no untagged protein	86 out of 1019 proteins (8.4%)[#]
Caulobacter ORFeome library [28,77]	mCherry RFP	*C. crescentus*	Chromosomal integration at specific locus (*xylX* gene), inducible promoter, endogenous gene untagged	333 out of 2786 proteins (12.0%)

[#] The authors quantified the amount of fluorescence intensity confined in the focus and called this measure "Spot%_protein", measured values are between 0 and 77.6%. 86 proteins have values above 5% and 269 proteins have values higher than 1%.

2.1.4. Validation of protein localization

Protein localization measurements are inherently difficult to validate. The perhaps most obvious validation approach is to measure spatial patterns in two separate experiments using independent methods, like immunofluorescence and FP fusions. Any discrepancy would then show that at least one method is flawed. This strategy was used, for example, to validate the localization of carboxysome fusions in *Synechococcus elongatus* PCC7942[78], septin ring fusions in *S. cerevisiae*[79,80] or FtsZ fusion proteins as discussed above.

However, both methods could also be flawed in the same way. This may sound unlikely, but since IF results are sensitive to conditions, it is not inconceivable that researchers convince themselves that the IF results that are consistent with FP are the correct ones,

and tone down conflicting results. To be conclusive, each assay must therefore first be subjected to thorough internal controls, and the final comparison should be performed in such a way as to exclude that one method is effectively adjusted to repeat the results of the other.

Many controls can be done to specifically test FP fusions. For example, one can show that the fluorescent protein used is spatially uniform when expressed alone and does not form any type of structure[81,82]. Further, one could find a mutation in the protein of interest that disrupts the observed localization[83] or determine the exact sequence that is necessary and sufficient for the localization of the fusion protein[23]. The most important test is perhaps for functionality because an FP fusion that disrupts the native localization is likely to also affect functionality. This is partly because any disruption suggests interference between the tag and the native protein, and partly because correct localization may be a prerequisite for correct function. Fluorescent protein fusions can easily be tested for functionality by complementation in the deletion background, if the deletion has a measureable phenotype, or by measuring activity. A fully functional fusion should behave identically or very similarly to the wild type whereas a non-functional fusion should show the deletion phenotype. Surprisingly, many fusions that are commonly used do not complement (or have not even been tested to determine whether they complement) and the wild-type allele is co-expressed to rescue the deletion phenotype. For example, all the FtsZ fluorescent protein fusions in *E. coli* are not fully functional and co-expression of untagged FtsZ is a must[41]. *C. crescentus* cells with a fluorescent protein fusion to the protease ClpP, which is essential in *Caulobacter*, are not viable and are hence only used in the presence of the wildtype allele as a second copy[28,29]. This approach of a second tagged copy was also used for the *E. coli* open reading frame archive (ASKA)[20] and the entire *C. crescentus* mCherry protein fusion library[28,77], suggesting that many fusions might not complement the deletion phenotype.

Non-functional fluorescent protein fusions are still relatively frequently utilized as "localization markers" but their usefulness is questionable.

Immunofluorescence is harder to evaluate with internal controls, but IF can be evaluated in conjunction with an FP fusion. For example, at least for non-essential proteins, the absence of false positives due to non-specific binding of the antibody can be established by performing the IF assay against the deletion strain, or under conditions where the target protein is not expressed. Specific binding can be tested by performing the IF against a strain that has the target protein tagged with an FP, and determining the degree of co-localization. Finally, the optimized IF protocol should be applied to wildtype cells (without the FP fusion) to determine whether the FP fusion interfered with localization. Though not even this procedure is 100% foolproof but it can effectively validate the IF assay even if the FP fusion interferes with localization.

2.1.5. The Clp proteases

The evolutionarily conserved[84] Clp proteases are a canonical example of bacterial proteins that were reported to form foci in several bacterial species[20,21,23-30] (**Figure 2.2**). The foci were suggested to play many important biological roles. It has been proposed for *Bacillus subtilis* (*B. subtilis*) that these foci reflect compartmentalized protein degradation factories, which localize close to the ribosomes to coordinate protein synthesis with protein degradation and quality control[24]. In *Mycobacterium tuberculosis* (*M. tuberculosis*), the ClpXP machinery was further suggested to control cell length[30], and in *C. crescentus*[28,29] foci formation of ClpXP was proposed to drive cell-cycle-dependent and spatially-localized degradation of key regulatory factors[29].

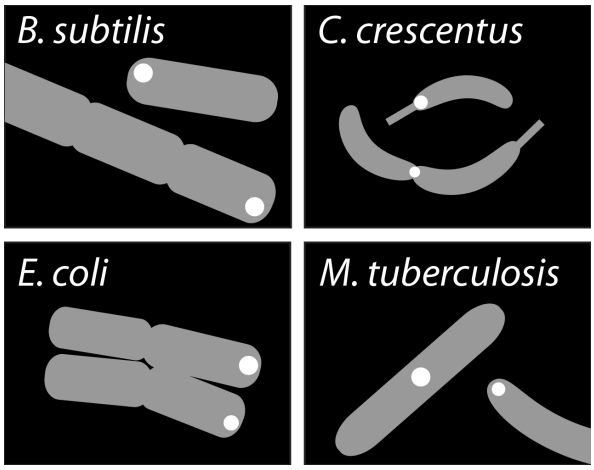


Figure 2.2. The bacterial Clp proteases localize in discrete foci in *B. subtilis*[23-27], *C. crescentus*[28,29], *E. coli*[20,21], *M. smegmatis*[30] and *M. tuberculosis*[30] cells. The foci preferentially localize to the cell poles and mid-cell region. Cells have usually one or zero foci.

The Clp proteases are barrel-shaped oligomeric complexes that are composed of four stacked rings. ClpP, a serine protease, forms the two inner rings, which are stacked head-to-head and constitute the proteolytic core. Each ClpP ring is composed of seven ClpP subunits. The ClpP cylinder is capped on either side by a hexameric chaperone ring, which is either formed by six ClpA or six ClpX molecules[85,86] (**Figure 2.3**).

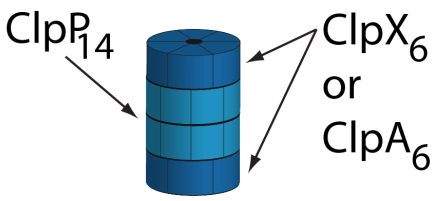


Figure 2.3. The Clp proteases are oligomeric complexes that are composed of four stacked rings. ClpP forms heptameric rings (light blue); ClpA and ClpX from hexameric rings (dark blue).

The ClpP 14-mer has two axial pores, which serve as the entry points for the unfolded substrate proteins. The ClpA and ClpX chaperone rings have ATPase activity and are involved in substrate recognition, substrate unfolding and substrate translocation to the proteolytic core where the substrates are hydrolyzed into short peptides[87,88].

The Clp protease complexes have many important physiological roles in the cell and degrade a diverse set of substrate proteins[88,89]. For example, the *E. coli* ClpAP and ClpXP proteases are involved in protein quality control. A damaged or truncated mRNA causes the translation machinery to stall. The stalled ribosome is then recognized by the tmRNA and binding of the tmRNA to the ribosome subsequently shifts the reading frame from the mRNA to the tmRNA. Translation of the tmRNA leads to the addition of the ssrA tag, a short eleven amino acid peptide tag (AANDENYALAA), to the nascent protein chain and marks the aberrant protein for degradation. The ssrA tagging system thus prevents the accumulation of incomplete proteins and frees stalled ribosomes[88-91]. Degradation of ssrA-tagged proteins is dependent upon ClpXP and to a smaller extent on ClpAP[92-94]. The adapter protein SspB further facilitates ClpXP-mediated degradation by enhancing substrate recognition[95] (**Figure 2.4a**).

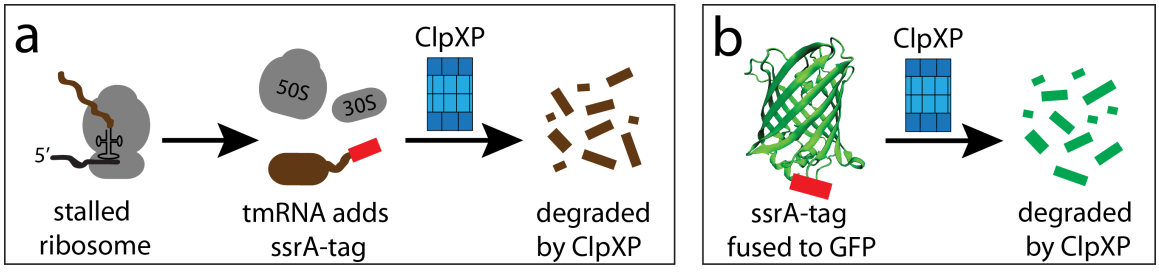


Figure 2.4. (**a**) The tmRNA frees stalled ribosomes and marks the aberrant polypeptide chain with an ssrA tag for destruction. (**b**) A genetic fusion between the ssrA tag and a stable protein leads to rapid degradation of the fusion protein as illustrated for GFP-ssrA.

The ssrA tag can also be used as a degradation tag to destabilize any protein of interest by genetically fusing the tag to the C-terminus of the target protein (**Figure 2.4b**). For example, ClpXP degrades GFP carrying a C-terminal ssrA tag (i.e. GFP-ssrA) rapidly *in vivo* with a half-life of less than one min as measured by pulse-chase experiments[93]. Higher half-lives were reported in cells by using inducible protein overexpression systems[96-104] but the higher values are probably due to protease saturation and fluorescent protein maturation issues since only the fraction of the synthesized GFP-ssrA molecules that live long enough to mature are measured. The GFP-ssrA fusion is commonly utilized as a model substrate for mechanistic studies of the Clp proteases *in vitro*[105,106]. Fusions to the ssrA tag are also frequently used in synthetic biology and single-cell imaging studies to generate more dynamic transcriptional reporters or to destabilize regulatory proteins of interest[96,98-104,107].

2.1.6. Our approach

The reported Clp protease foci display large cell-to-cell variability, wherein cells can have one, two or zero foci. This seems surprising for proteases that are responsible for the degradation of many important substrates that are involved in central pathways, since it could broadly randomize protein abundances or even temporarily abolish degradation in cells with zero foci. This prompted us to study these processes more quantitatively in single cells and to test whether the previously reported foci of the Clp proteases reflect the native protein localization or are artifactual of the methods used. In fact, though Clp foci have been reported by numerous different labs, using different fluorescent proteins and different types of bacteria, all studies are based on FP fusions without any independent validation. Because the FPs used generally do not form foci when expressed alone, it seemed far-fetched that the results would be false, but on the

other hand it also seemed far-fetched that cells would abolish central proteolysis for a significant part of the cell cycle if the focus segregated to the other daughter.

Because we initially did not have access to high-quality antibodies, our first priority was to design a conclusive and non-intrusive alternative validation method. Our approaches independently prove that the observed foci were artifacts of the FPs used. We then proceeded to evaluate a large number of different FPs, attempted to explain the result mechanistically, and showed how widespread these types of artifacts are in other systems. Though we, of course, know that bacteria are not just bags on enzymes, in all the systems we have analyzed so far, the native proteins were close to being spatially homogenous and the distinct patterns reported were all artifacts of the FPs.

2.2. Results

2.2.1. Clp protease foci show binary segregation at cell division

We first fused two commonly used fluorescent proteins, Venus and superfolder GFP (sfGFP), to the ClpX and ClpP proteins[†] in *E. coli* and confirmed the previously reported localization patterns, typically observing a single bright focus in roughly half of the cells during balanced exponential growth in rich medium. Cells without a focus showed low cytoplasmic fluorescence, similar to the cytoplasmic signal in foci-harboring cells. Tracking cells through division showed how the focus segregated to one of the two daughters (**Figure 2.5**), whereas the other cell formed a new focus within a few generations (**Video 2.1**).

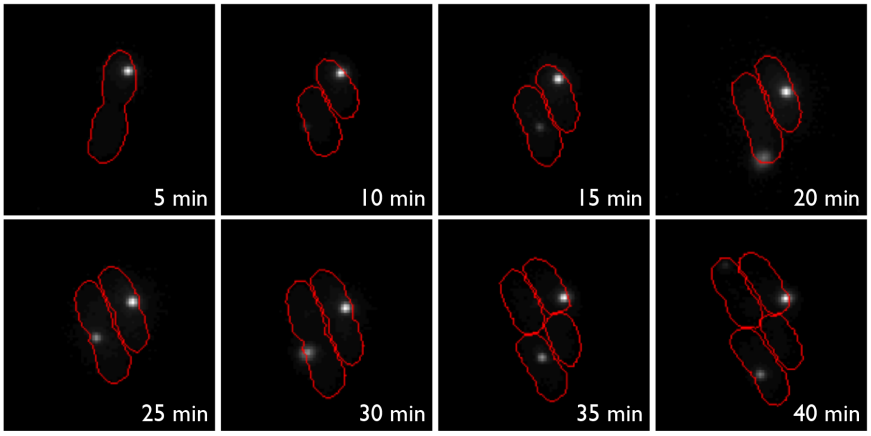


Figure 2.5. The ClpP-Venus YFP fusion localizes into bright fluorescent foci. Cells have either one or zero foci. At cell division, the foci show binary segregation and only one daughter inherits the focus. The cells also have a weak cytoplasmic YFP signal, which is not visible here because of the bright foci.

[†] All fusions were constructed at the endogenous gene loci and the tagged protein is the only type of that protein in the cell. Strains that harbor the fusion proteins do not also contain the wild-type allele.

Cell division creates two daughter cells with apparently different fates in respect to protein degradation. The observed localization of the fluorescent protein-labeled Clp protease foci should thus cause substantial single-cell heterogeneity after division in the turnover rates of protease substrates.

2.2.2. Single-cell segregation analysis can be used to test protein localization patterns in live cells

To investigate whether the *E. coli* ClpX, ClpA and ClpP oligomers truly form foci in living cells, without relying on additional methods to visualize the Clp complexes, we designed a novel single-cell, function-based segregation assay. On a conceptual level, our approach exploits the fact that localization patterns determine the statistical differences between the two daughter cells right after cell division and thereby influence the heterogeneity after division in any affected downstream processes. By measuring the downstream heterogeneity in the presence and absence of a fluorescent tag to the protein of interest, a side-by-side comparison should reveal whether the tag interferes with protein localization. For non-intrusive tags, the single-cell heterogeneity after division in the downstream process should be the same with or without the tag to the protein of interest (**Figure 2.6**).

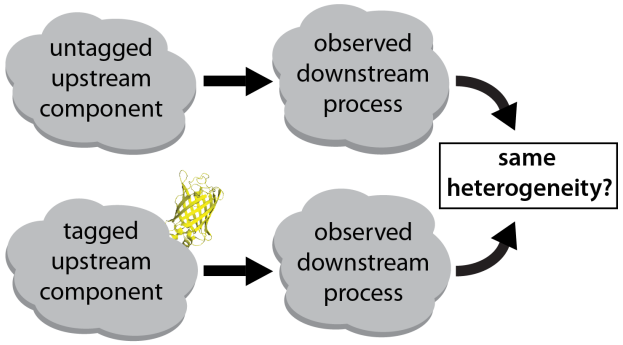


Figure 2.6. Schematic of the segregation assay. An upstream process (e.g. Clp protease localization) affects a downstream process (e.g. substrate degradation) that can be measured in daughter cells originating from cells with and without a tag on the upstream component. If the tag is nonintrusive, the heterogeneity of the downstream process in the daughter cells should be tag-independent.

2.2.3. Segregation analysis reveals large heterogeneity in substrate degradation rates in the strains with ClpP foci but no variability in the wild type

We observed that the localization of the Clp proteases in foci causes substantial single-cell heterogeneity in the Clp protein levels after cell division, when fused to the two FPs used. One daughter cell thus inherits the focus while the other receives very few protease molecules. This protease cell-to-cell variability should in turn create heterogeneity in the turnover rates of substrates degraded by the Clp proteases.

We then used dual-color time-lapse microscopy to simultaneously measure substrate abundances and protease localization patterns in individual *E. coli* cells over time. As the reporter substrate, we fused mCherry[70] to the *E. coli* ssrA tag, which marks mCherry for proteolysis by ClpXP (**Figure 2.7**) and to a lesser extent by ClpAP[92-94].

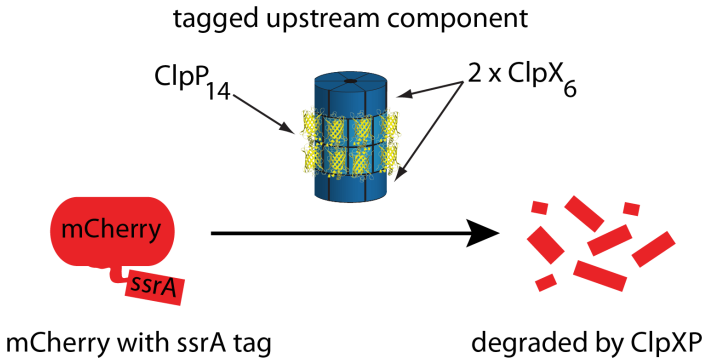


Figure 2.7. Schematic of the mCherry-ssrA degradation reporter system. The upstream component is the protease, which is tagged with Venus. The downstream process that is monitored is the degradation of the mCherry-ssrA substrate by the Clp proteases.

We expressed mCherry-ssrA to high levels from an inducible promoter in both foci-forming fluorescent protein strains and in the wild-type strain under identical conditions, washed away the inducer, waited for the fluorophores to mature, and then followed the fate of the mCherry-ssrA degradation reporter over time. We measured the reporter degradation rate in daughter cells over one cell cycle after division and analyzed the heterogeneity between individual daughter cell pairs. The observed variability in protein degradation rates between siblings should thus be the same in the presence and absence of an FP fused to the protease, unless the FP tag interferes with protein localization and segregation. Note that this test does not rely on the mCherry reporter being non-intrusive, or on any indirect arguments based on assumptions about the underlying biology. It is a direct comparison of a downstream process with and without a tag on an upstream process, and only tests for equivalence.

When we tagged either ClpX or ClpP with Venus or sfGFP, cells that contained the focus actively degraded mCherry-ssrA, whereas cells without a focus showed mild to extreme reduction in mCherry-ssrA degradation, thus producing two daughters with very different

mCherry-ssrA degradation rates (**Figure 2.8**, **Figure 2.9a** and **Video 2.2**). However, both daughter cells in the wild-type strain, without an FP fused to the protease, continued proteolysis of mCherry-ssrA at very similar rates (**Figure 2.8**, **Figure 2.9b** and **Video 2.3**). This shows that the fluorescent protein tag causes severe artifacts and that ClpX and ClpP fluorescent protein fusions cannot be trusted for determining localization of native, untagged proteins.

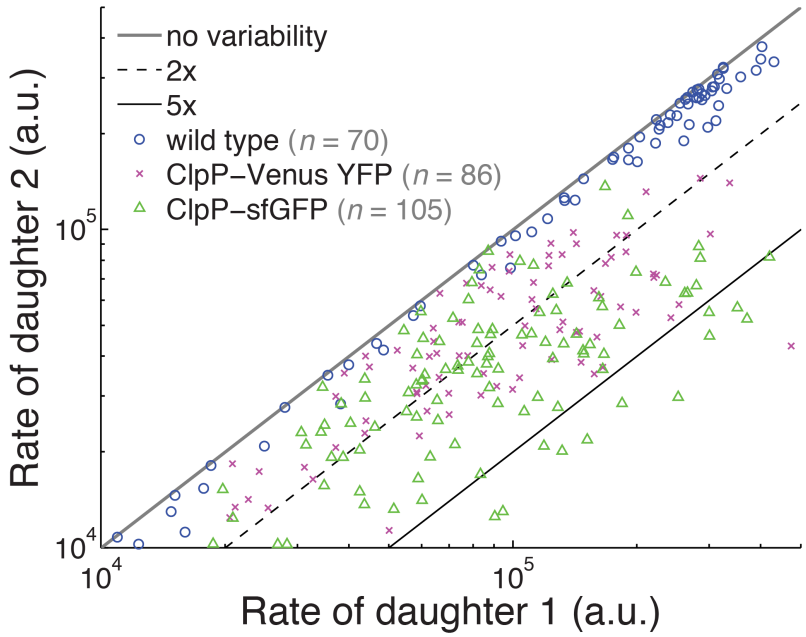


Figure 2.8. Single-cell segregation assay reveals that ClpP foci are artifactual. Single-cell mCherry-ssrA degradation rates in arbitrary units (a.u.) measured by monitoring total RFP fluorescence intensity per cell in daughter cells over time after cell division in indicated bacterial strains. The daughter cell with the faster degradation rate was designated cell 1. The spread along the diagonal is due to the pulse induction of the mCherry-ssrA reporter and does not reflect in any way a cell-to-cell variability in protease levels or activity. Diagonal lines represent no cell-to-cell variability (gray line), 2x variability (dashed black line) and 5x variability (solid black line)

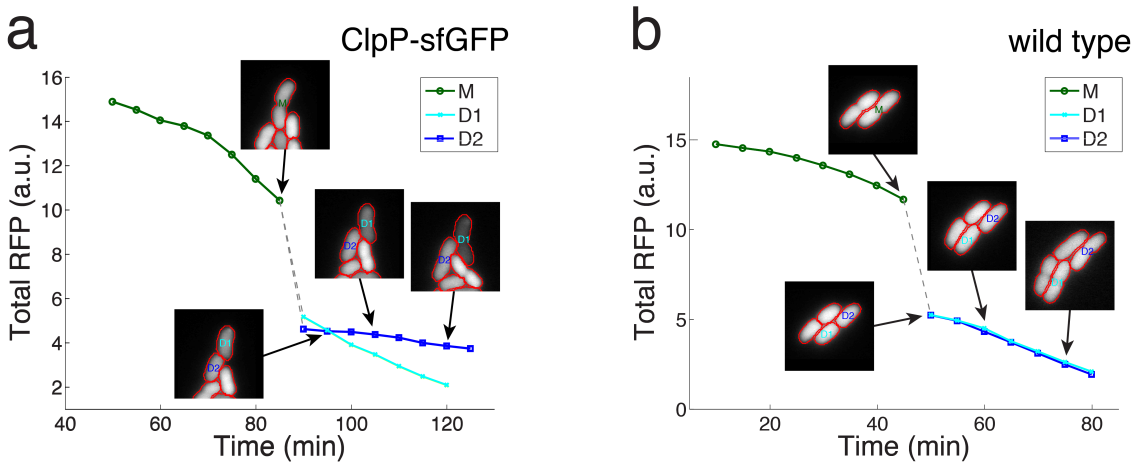


Figure 2.9. ClpP-fluorescent protein foci generate cell-to-cell variability after cell division. (**a**) A mother cell (M, green) with a ClpP-sfGFP foci (strain DHL955) divides and daughter cell 1 (D1, cyan) inherits the ClpP-sfGFP focus (GFP images are not shown) and degrades an mCherry-ssrA reporter substrate much faster than daughter cell 2 (D2, blue). RFP images (insets) are shown for selected time-points. Cell outlines in red. The degradation rate ratio (defined as degradation rate of the daughter with the faster rate divided by the daughter with the slower rate) is 4.02. (**b**) In strong contrast, very little cell-to-cell variability after division was observed for the wildtype (DHL440). Both daughters (D1 and D2) have a very similar degradation rate after cell division. The degradation rate ratio is 1.04 for this sibling pair of the wild-type strain. The respective time-lapse movies were recorded at 30 ˚C (similar results at 37 ˚C). The mCherry-ssrA degradation reporter is degraded in single cells with zeroth-order kinetics, as previously shown for GFP-ssrA[103]. The deviation of the RFP curve from a straight line at the beginning of the time-lapse experiment is probably because of delayed mCherry maturation and competition of non-matured mCherry molecules for degradation.

2.2.4. Immunofluorescence reveals uniformly distributed ClpX particles

To further validate our results, we performed several independent tests. Immunofluorescence microscopy with antibodies to ClpX (**Figure 2.10**) in the strain expressing the foci-forming ClpX-Venus as a positive control and a *clpX* knockout strain as a negative control, confirmed that ClpX antibodies were specific and that fixation did not disassemble the ClpX-Venus foci. For wild-type cells, the immunofluorescence images have near single-particle resolution and indicated that ClpX formed 20–50 complexes that were uniformly distributed in the cell.

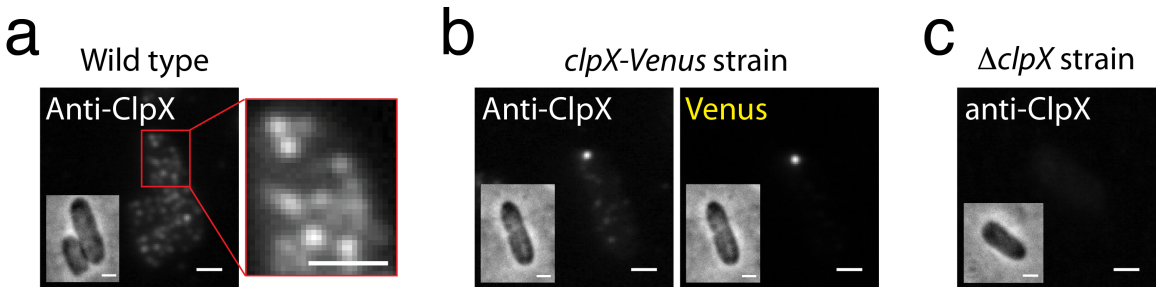


Figure 2.10. Immunofluorescence microscopy with an antibody against the ClpX protein (anti-ClpX). (**a**) Wild-type cells. Magnification of the boxed region is shown on the right, (**b**) *clpX*-Venus strain. Venus fluorescence is shown on the right. (**c**) Negative control is the Δ*clpX* strain. Insets are phase-contrast images. The scale bars (white) are 1 µm.

2.2.5. SNAP tag and HaloTag fusions to ClpP reveal individual ClpP particles

I further constructed and analyzed ClpP fusions to the monomeric SNAP tag (**Figure 2.11a**) and HaloTag (**Figure 2.11b**) and observed a uniform spatial distribution of the ClpP particles with no signs of foci formation. This provides further independent confirmation of the main result, but also provides an alternative method for monitoring the actual behavior of the Clp complexes. The SNAP tag is derived from a human DNA repair enzyme[108] and the HaloTag from a bacterial haloalkane dehalogenase[109-111]; both tags are unrelated to any fluorescent protein. The results achieved with the SNAP tag and HaloTag showed the same uniform spatial distribution and perfectly agree with the immunofluorescence and the single-cell segregation assay of the wild-type strain.

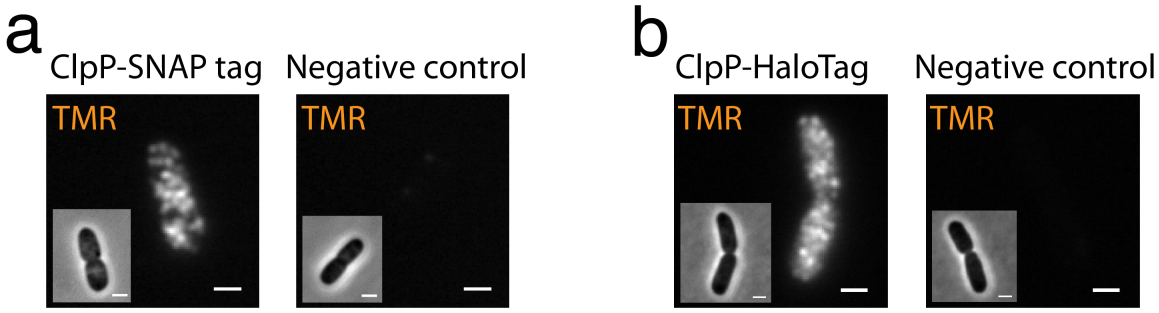


Figure 2.11. Fluorescence images of bacteria expressing (**a**) ClpP-SNAP tag or (**b**) ClpP-HaloTag fusion proteins labeled with the organic dye Tetramethylrhodamine (TMR), compared to labeling of wild-type cells that lack the SNAP tag or HaloTag proteins (negative control). Insets are phase images. The scale bars (white) correspond to 1 μm.

2.2.6. Observed protein localization depends on the fluorescent protein used

These findings motivated us to evaluate ~20 different fluorescent proteins (**Figure 2.12** and **Table 2.6**) for foci formation when fused to ClpP or ClpX.

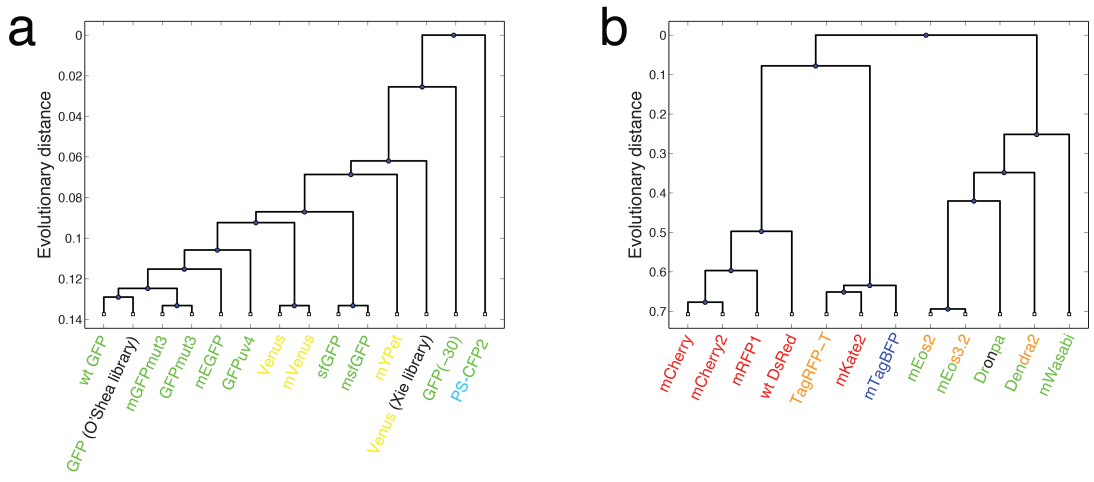


Figure 2.12. Phylogenetic tree of (**a**) GFP-like and (**b**) DsRed-like fluorescent proteins. All proteins except wild-type (wt) GFP, GFP (O'Shea library[58]), Venus (Xie library[21]), mRFP1, wt DsRed, mRFP1 and mEos3.2 were tested.

We found that sfGFP[112], GFPuv4[113], Venus[114], mCherry[70] and mCherry2 all caused substantial foci formation, despite being monomers or very weak dimers when expressed alone. mKate2[115] and TagRFP-T[116] caused intermediate clustering, whereas with mVenus[117] and mYPet[118] most of the fluorescence signal was spatially uniform, although we observed foci in a few cells. mTagBFP[119], mWasabi[120] and mEos2[121] fusions resulted in a weak signal with infrequent dim foci. We detected no foci for photo-switchable (PS)-CFP2[122], reversible switchable (rs)FastLime[123] (data not shown) and a mutant of sfGFP with a charge of −30 (GFP(−30))[124], but the signals were very dim. Finally, mGFPmut3, msfGFP[125], mEGFP[126], Dronpa[127] and Dendra2[128] displayed an essentially uniform signal. Fluorescent protein fusions to ClpP generally caused more foci formation than fusions to ClpX, in particular for mYPet (**Figure 2.13**).

Figure 2.13

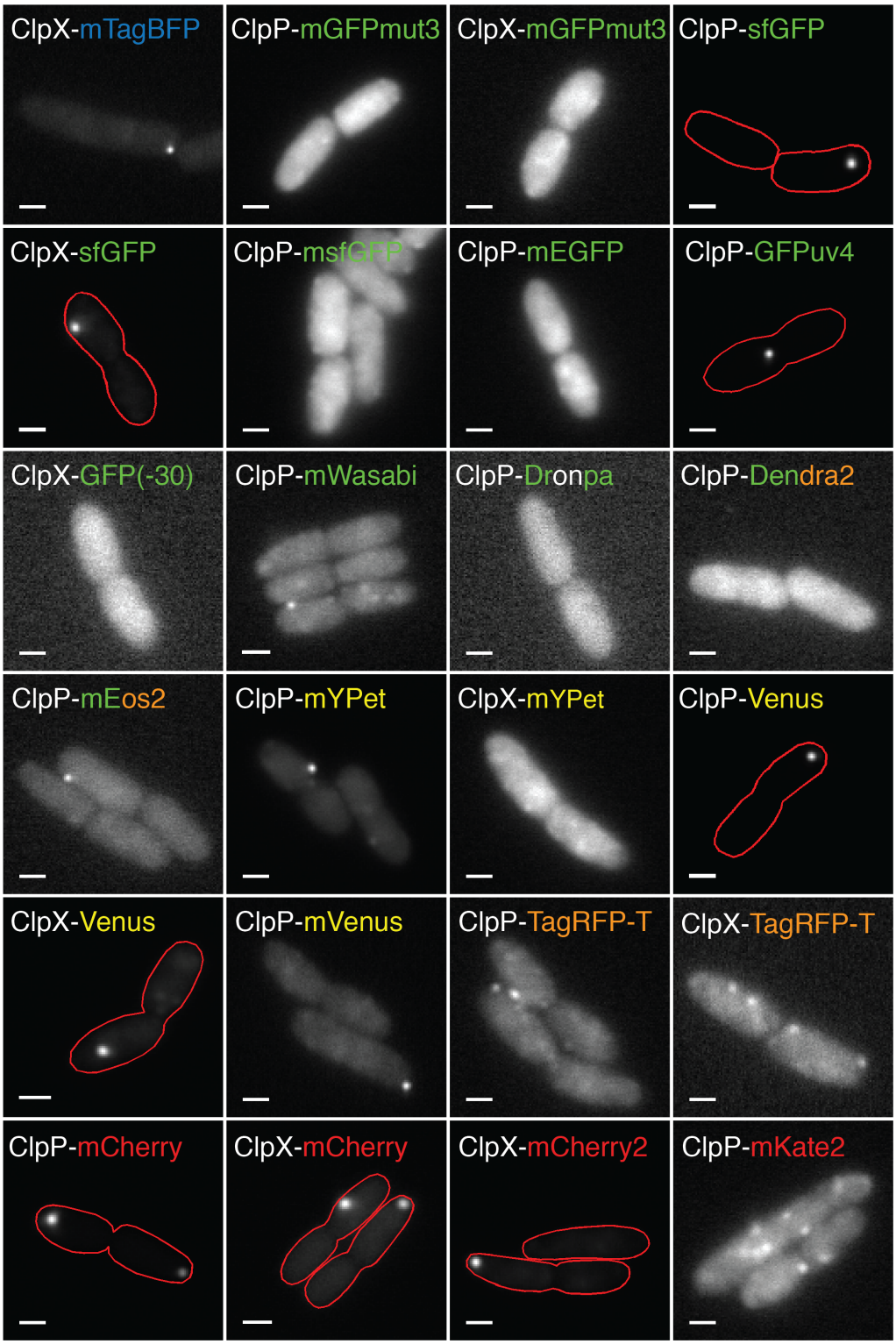

ClpX-mTagBFP
ClpP-mGFPmut3
ClpX-mGFPmut3
ClpP-sfGFP
ClpX-sfGFP
ClpP-msfGFP
ClpP-mEGFP
ClpP-GFPuv4
ClpX-GFP(-30)
ClpP-mWasabi
ClpP-Dronpa
ClpP-Dendra2
ClpP-mEos2
ClpP-mYPet
ClpX-mYPet
ClpP-Venus
ClpX-Venus
ClpP-mVenus
ClpP-TagRFP-T
ClpX-TagRFP-T
ClpP-mCherry
ClpX-mCherry
ClpX-mCherry2
ClpP-mKate2

Figure 2.13 (Continued).

Figure 2.13. Fluorescence images of bacterial cells expressing the indicated constructs. The ClpP-FP and ClpX-FP fusions form a single discrete focus (e.g. Venus YFP), cytoplasmic puncta (e.g. TagRFP-T) or give rise to a relatively uniform cytoplasmic stain (e.g. mGFPmut3). We detected no foci for ClpX-PS-CFP2, ClpX-Dronpa, ClpX-rsFastLime (all data not shown) and ClpX-GFP(–30) but the signal was very dim. The foci observed with ClpP-mYPet and ClpP-mVenus YFP were less frequent and contained less of the total fluorescence probably (compared to ClpP-Venus YFP) because of the monomeric A206K[31] mutation. The cell outline (red) is shown for cells with weak cytoplasmic signal. The scale bars (white) are 1 μm.

Foci-forming tendencies could also be affected by protein expression amounts and protein levels, which in turn could be affected by the fluorescent protein tag. To test this, we expressed two separate copies of the gene for ClpP-mGFPmut3 in the same strain. We observed no increase in clustering despite doubling the expression level (**Figure 2.14**).

Figure 2.14

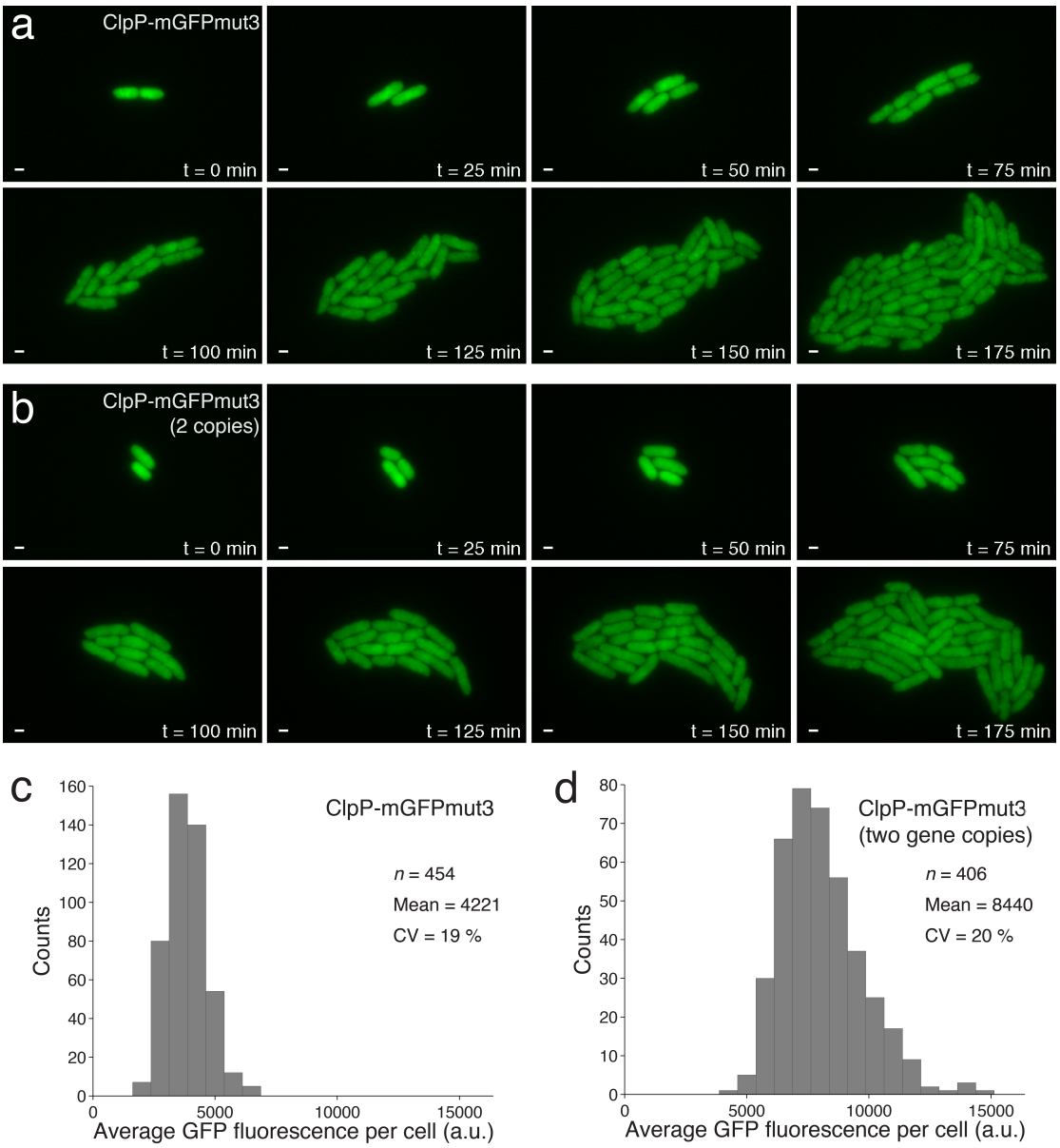


Figure 2.14 (Continued).

Figure 2.14. Micro-colony growth of *E. coli* cells with (**a**) one copy of ClpP-mGFPmut3 (strain DHL747) and (**b**) two copies of the ClpP-mGFPmut3 fusion (strain DHL987). The first copy corresponds to the endogenous *clpP* gene tagged with mGFPmut3. The second ClpP-mGFPmut3 copy was integrated in the *att*Tn7 site with expression controlled by the native *clpP* promoter. mCherry-ssrA was pulse-induced with IPTG for ~2 h and cells were spotted on an agar pad after washing away the IPTG. GFP images (green) were acquired every 25 min and RFP images every 5 min (data not shown). Cell growth and imaging was carried out at 30 ℃. The ClpP-mGFPmut3 fusions show a relatively uniform distribution, although occasionally a weak punctate localization with dim foci can be seen (probably because mGFPmut3 is still not truly monomeric or some ClpP molecules are less mobile or even interact with the cell membrane). Bright fluorescent foci (as seen for e.g. with ClpP-Venus, ClpP-sfGFP or ClpP- mCherry) were not observed. The scale bars (white) correspond to 1 μm. The average GFP fluorescence per cell is approximately 2.0x as high for (**d**) the strain with two copies of ClpP-mGFPmut3 as for (**c**) the strain with just the endogenous *clpP* tagged with mGFPmut3. Data was pooled from 10 micro-colonies for each strain. The GFP levels are corrected for cellular autofluorescence by imaging the MC4100 wild-type strain with the same GFP exposure time. The average GFP fluorescence per cell is in arbitrary units (a.u.).

Western blot analysis of a few different tags (Venus, sfGFP, mGFPmut3 and the SNAP tag) fused to ClpP shows that the ClpP fusion protein levels vary for different tags probably because the tags have different effects on translation (different codon usage), protein folding, incorporation in the complex and protein stability (**Figure 2.15a**). ClpP-mGFPmut3 levels were reduced at 37 ˚C but were similar to the wild type at 30 ˚C. The strain that carries two *clpP*-mGFPmut3 gene copies has levels that are higher than the wild type (**Figure 2.15b**). A few ClpP fusions also exert polarity effects on the downstream *clpX* gene. For example, the ClpP-Venus fusion increases expression of the *clpX* gene whereas no effect was observed for the ClpP-sfGFP fusion (**Figure 2.15c**).

Figure 2.15

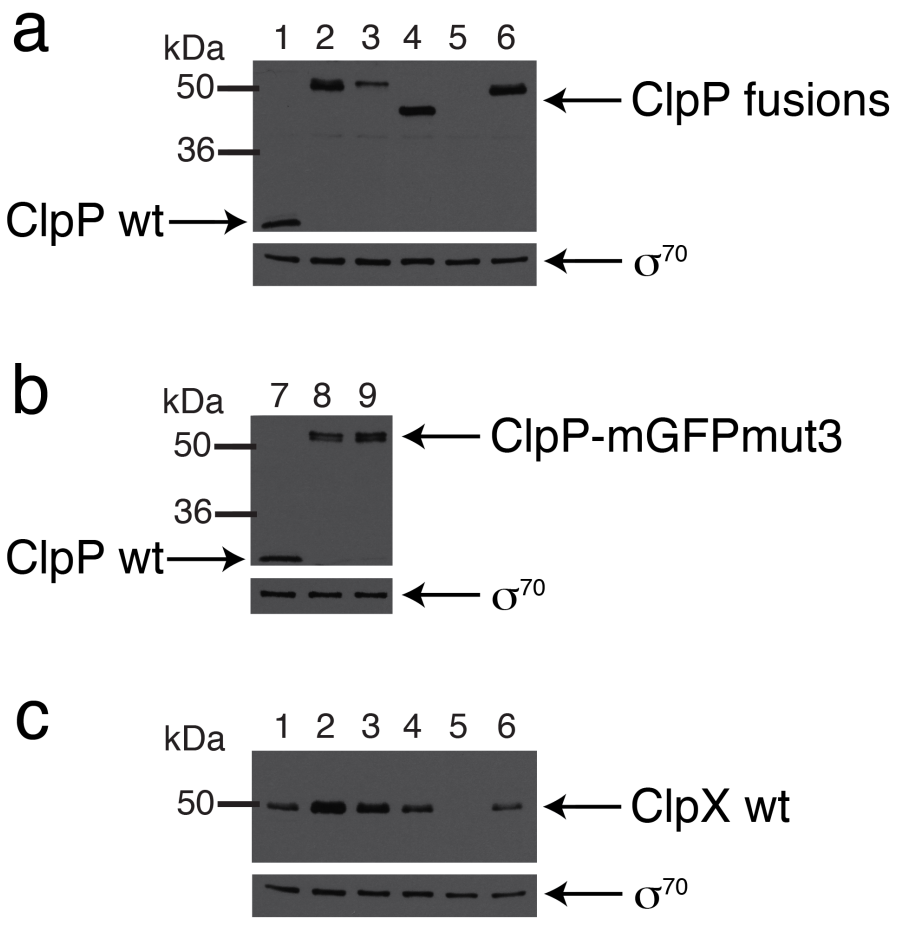


Figure 2.15 (Continued).

Figure 2.15. Western blot analysis shows that the ClpP-fluorescent protein levels vary for different tags. (**a**) The levels of the ClpP-Venus (lane 2: strain DHL436), ClpP-SNAP tag (lane 4: strain DHL663) and ClpP-sfGFP (lane 6: strain DHL778) fusions are higher than the wild type (lane 1: MC4100) whereas the ClpP-mGFPmut3 (lane 3: strain DHL661) fusion is lower than the wild type (wt). Samples were prepared from exponential-phase cells grown at 37 ℃ and probed with the anti-ClpP antibody. The level of the ClpP-mGFPmut3 fusion is reduced, probably because mGFPmut3 lacks the F64L mutation, which increases the folding efficiency at 37℃[129]. (**b**) Samples were prepared from exponential-phase cells grown at 30 ℃ and probed with the anti-ClpP antibody. The ClpP-mGFPmut3 level (lane 8: strain DHL661) is similar to the wild type (lane 7: strain MC4100) at 30 ℃. A strain expressing two copies of ClpP- mGFPmut3 has higher ClpP levels (lane 9: strain DHL986) than the wild type at 30 ℃ (lane 7), and the average GFP fluorescence per cell is ~2.0x as high as for the strain with one copy of ClpP-mGFPmut3 (**Figure 2.14**). The ClpP-mGFPmut3 band doublet (lane 8 and 9) corresponds probably to ClpP-mGFPmut3 with and without the 14 amino acid leader sequence[130]. (**c**) The ClpX abundance is also affected by the various fluorescent protein tags to ClpP probably due to polarity effects. Western blot samples were prepared from exponential-phase cells grown at 37 ℃ and probed with the anti-ClpX antibody. The anti-ClpP and anti-ClpX antibodies are specific for their respective antigens, as confirmed with an Δ*clpPX* deletion strain (lanes 5: strain DHL708). σ^{70} (also known as RpoD) levels were measured and used as a loading control.

2.2.7. The Clp proteases are uniformly distributed inside *E. coli* cells and are not localized in foci

We further used highly inclined and laminated optical sheet (HILO) microscopy[131] and imaged fixed cells and live cells in real time. Gentle fixation of cells expressing ClpX-mGFPmut3 and ClpP-mGFPmut3 again revealed uniformly distributed complexes, resembling the immunofluorescence result (**Figure 2.16**).

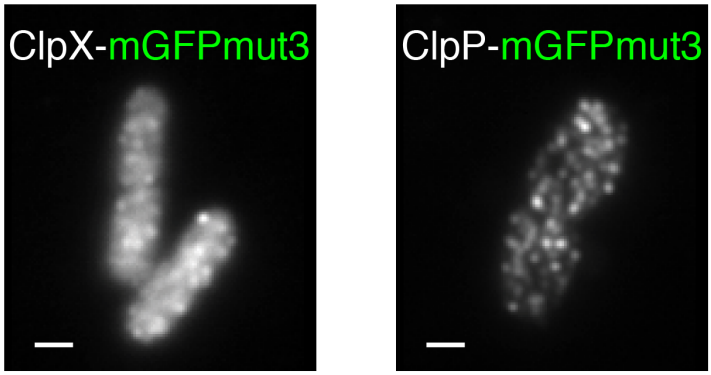


Figure 2.16. HILO microscopy of fixed *E. coli* cells harboring the ClpX-mGFPmut3 (left) or the ClpP-mGFPmut3 (right) fusion. Scale bars (white) are 1 μm.

In live cells, we observed ClpA-mGFPmut3, ClpP-mGFPmut3 and ClpX-mGFPmut3 complexes moving freely and rapidly, occupying all of the cytoplasm without significantly favoring any particular region (**Videos 2.4**–**2.7** and **Figure 2.17**).

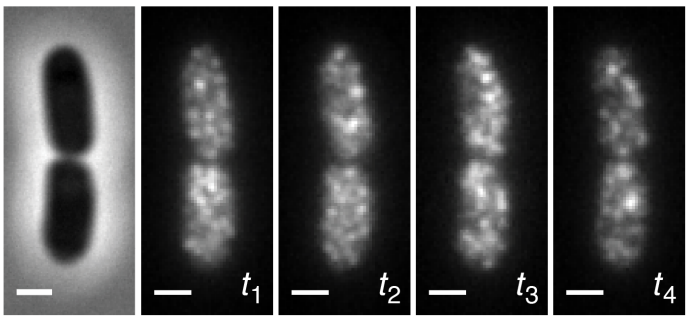


Figure 2.17. Phase-contrast (left) and live-cell HILO microscopy fluorescence images (right) of *E. coli* cells expressing ClpP-mGFPmut3 taken one second apart at time points t_1–t_4. Scale bars correspond to 1 μm.

We also detected individual ClpP-Dronpa molecules in live cells with HILO imaging and these molecules were also uniformly distributed (**Figure 2.18**).

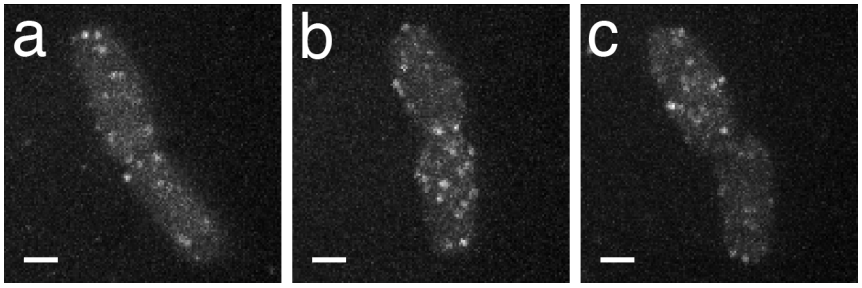


Figure 2.18. ClpP-Dronpa molecules are uniformly distributed inside live *E. coli* cells. Individual *E. coli* cells expressing the ClpP-Dronpa fusion (strain DHL726) were imaged under constant illumination with a 488 nm laser in HILO mode and a time-series was recorded. Dronpa molecules were not photo-converted with a 405 nm laser during the experiment. The maximum projection is shown for three different time-series (**a**–**c**). The scale bars (white) correspond to 1 μm.

The measured dynamics of the protease particles further supports the segregation assay and the localization patterns observed with the methods described above, and makes evolutionary sense in that free diffusion would ensure rapid degradation of substrate proteins even if any substrates were localized.

2.2.8. Segregation analysis reveals very little heterogeneity in substrate degradation rates in daughters of strains with the non-foci-forming tags

We then analyzed the ClpP-SNAP tag, ClpP-mGFPmut3, ClpP-Dronpa and ClpP-Dendra2 fusions using our assay and observed very little cell-to-cell variability after division, confirming that these tags, though not perfectly mimicking the wild type, were less prone to clustering artifacts (**Figure 2.19**). We imaged the ClpP-mGFPmut3 strains and the wild type at 30 °C because ClpP-mGFPmut3 amounts were reduced at 37 °C but were similar to those of the wild type at 30 °C (**Figures 2.14**, **2.15a** and **2.15b**). All ClpP fluorescent protein fusions also showed a somewhat reduced degradation activity compared to the wild type, presumably because the bulky fluorescent protein tags interfered with protease activity. Of all the reporters tested, the SNAP tag was both the most active and least intrusive in terms of localization.

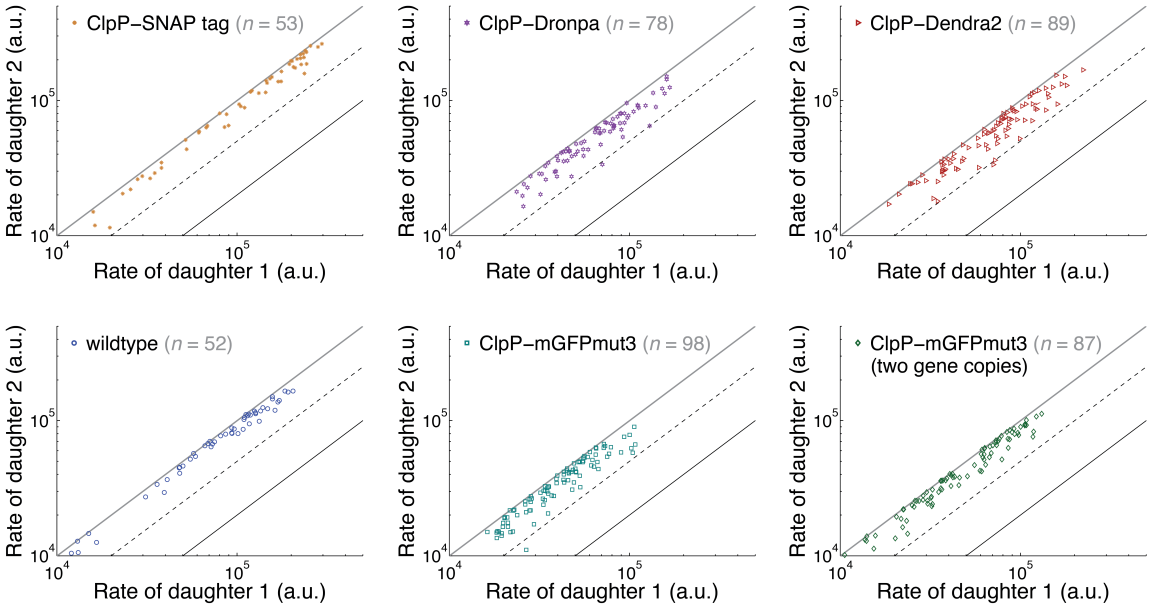


Figure 2.19. Single-cell segregation assays of the non-foci-forming tags. Degradation rates of the mCherry-ssrA reporter in arbitrary units (a.u.), measured by monitoring total red fluorescence intensity per cell in daughter cell pairs expressing the indicated proteins by time-lapse microscopy at 37 °C (upper row) and 30 °C (lower row). Diagonal lines represent no cell-to-cell variability (gray), 2x variability (dashed black) and 5x variability (solid black).

2.2.9. The Clp-FP foci are not protein aggregates

The terms aggregation and oligomerization are sometimes used as if they were synonyms when describing problems with fluorescent protein fusions. Aggregation of a fluorescent protein fusion is usually due to folding problems of the fusion (or of the fluorescent protein alone) and results in a non-functional fusion protein that forms cytoplasmic aggregates (also known as inclusion bodies), which are visible in the phase image. This was not the case for the Clp foci, which were only observed in the fluorescent images. Oligomerization of fluorescent proteins is not due to protein folding problems but because of a high binding affinity of the fusions to each other (i.e. a self-association tendency). Fluorescent protein oligomerization will increase the molecular weight of the fusion, decrease the diffusion constant and could cause mislocalization artifacts.

The bright ClpP-fluorescent protein foci were proteolytically active and highly fluorescent, showing that the fusions were functional and not misfolded. Introducing the monomeric A206K mutation[31] into superfolder GFP or Venus decreases the dimerization tendency and either abolished or substantially reduced foci formation (**Figure 2.13**), again demonstrating that the foci were not caused by spontaneously misfolded fluorescent proteins. However, foci formation cannot solely be attributed to the FPs themselves; even the strongest foci-forming fluorescent proteins, such as superfolder GFP or Venus, yielded spatially uniform fluorescence when expressed alone, even at high levels (**Figure 2.20**).

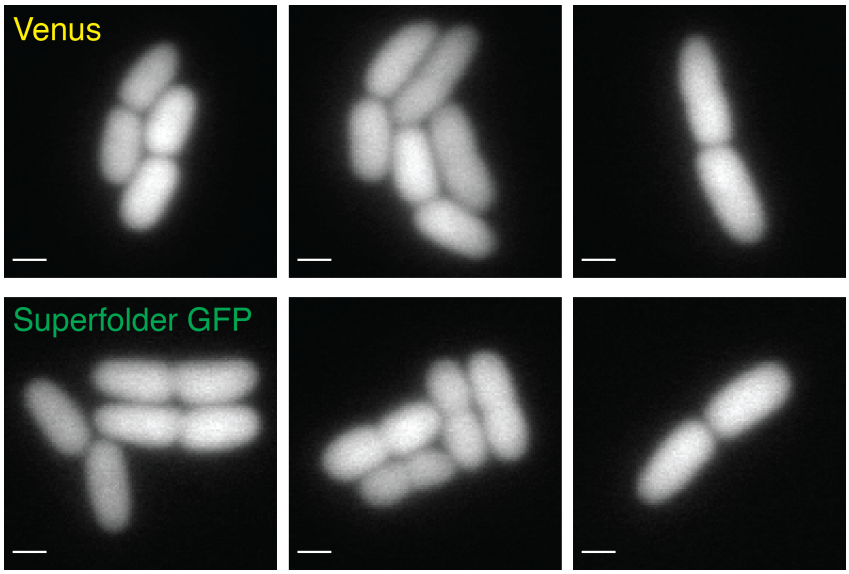


Figure 2.20. Micrographs of *E. coli* cells expressing Venus (upper row: strain DHL362) or superfolder GFP (lower row: strain DHL409) alone. The fluorescent proteins are expressed from the strong P_{LlacO1} promoter on a pSC101 plasmid. Fluorescent protein foci or aggregates do not form despite high intracellular levels. The scale bars (white) correspond to 1 μm.

2.2.10. Foci formation is likely due to avidity effects

Expression of both the Clp complexes and the fluorescent protein is thus spatially uniform when each is expressed alone, and they only form foci when fused to each other. We hypothesize that this is due to avidity effects. In fluorescent protein fusions, the homo-oligomeric proteins could act as scaffolds, bringing several fluorescent proteins in close proximity to each other. This would prevent the oligomers from diffusing apart after a single fluorescent protein-fluorescent protein dissociation event and allow them to rebind before the remaining links are broken (**Figure 2.21**), thereby driving the coalescence of tagged oligomers into visible foci. This phenomenon is known as avidity and was first described for antibodies that can bind multiple-epitope antigens with effective affinities that are orders of magnitude stronger than the single ligand-binding site alone[132].

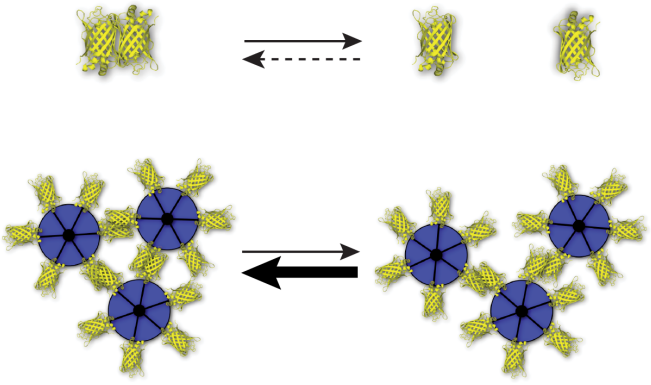


Figure 2.21. Schematic of antiparallel fluorescent protein dimer (yellow) dissociating into monomers (top) and of avidity effects potentially clustering tagged ClpX hexamers (blue; bottom). For simple monomers, re-association is slow because the molecules diffuse apart; for tagged oligomers, the complex acts as a scaffold to facilitate re-association.

2.2.11. Artifactual foci formation is widespread

These results raise the question of how many other reported foci are caused or greatly exaggerated by the fluorescent protein fusions. In fact, the fluorescent proteins we observed to be prone to clustering are used in the three main bacterial fluorescent protein fusion libraries, mCherry in a Caulobacter crescentus library[28,77], Venus in an *E. coli* library[21] and GFPuv4 in the *E. coli* open reading frame archive (ASKA)[20], that all report many foci. The ClpX and ClpP foci have even been used as positive controls in a genome-wide localization study[28,77]. To investigate whether fluorescent proteins caused false protein localization patterns more generally, we used one of the fluorescent proteins we found to be very monomeric, mGFPmut3, to retag five additional *E. coli* proteins (Hfq, PepP, IbpA, FruK and MviM) that had been previously reported to form bright foci in multiple fluorescent protein libraries[20,21] (**Table 2.2**).

Table 2.2

Protein name	Protein function	Oligomerization state	ASKA library[20]	Venus library[21] (Spot%_Protein)	mGFPmut3 fusions (this study)
Hfq	RNA chaperone	Hexamer[133]	focus	64.27	mainly uniform with occasional dim foci
PepP	Aminopeptidase	tetramer[134,135]	focus	20.317	uniform
IbpA	small heat shock protein	polydisperse oligomers (100–150 subunits)[136]; fibrils *in vitro*[137]	focus	11.85	mainly uniform with some punctate foci
MviM	putative oxidoreductase	dimer* (PDB: 1TLT)	focus	8.828	uniform
FruK	fructose-1-phosphate kinase	dimer*[138]	focus	77.586	uniform

* The *in vivo* oligomerization status is not well studied.

The fusions PepP-mGFPmut3, FruK-mGFPmut3 and MviM-mGFPmut3 showed no foci, whereas Hfq-mGFPmut3 and IbpA-mGFPmut3 were uniformly distributed in most cells and only showed dim foci in a small fraction of cells (**Figure 2.22** and **Video 2.8**). The Hfq-mGFPmut3 fusion did not form foci but was found throughout the entire cell. This localization pattern is in agreement with published immunofluorescence[139] and electron microscopy[140] data. An occasional punctate localization with dim foci was observed (magnified cell in **Figure 2.22a**) though even the brightest of the dim foci has less than 12% of the fluorescence intensity confined in the focus (data not shown), which is in strong contrast to a previously reported[21] 64% with an Hfq-Venus fusion. The IbpA-mGFPmut3 fusion displayed a punctate localization pattern but bright single foci, as previously reported[20], were not observed (**Figure 2.22c**). The small heat shock protein IbpA can form large polydisperse oligomers[136,137], which is in agreement with my observed localization pattern. Bright fluorescent foci, resembling e.g. the ClpP-Venus foci, were not observed for any of the tested fusion proteins.

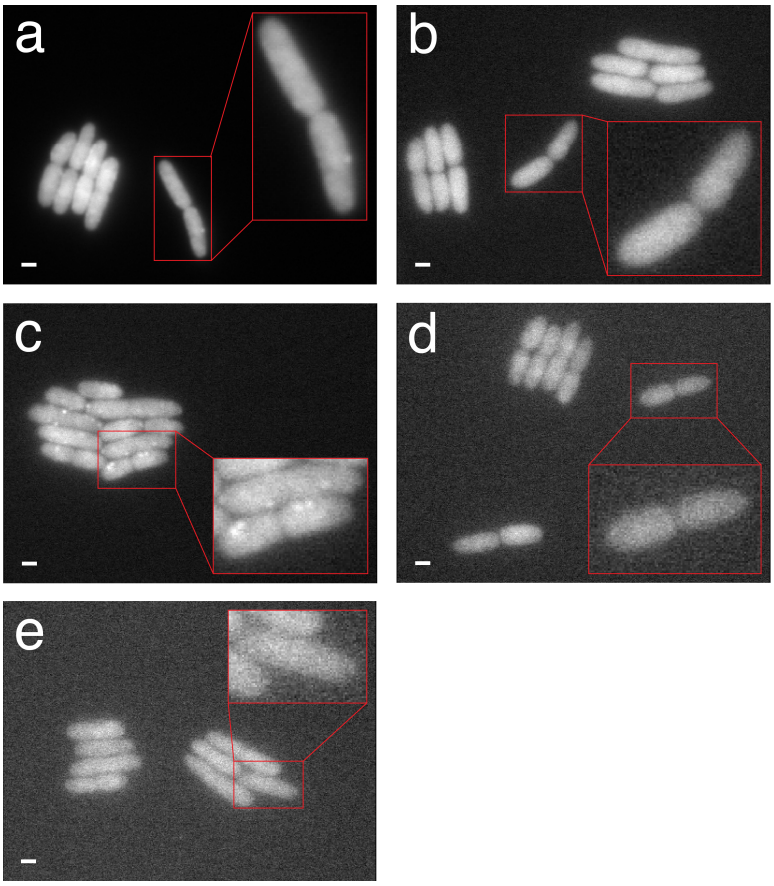


Figure 2.22. False protein localization, mediated by clustering of fluorescent proteins, seems to be a widespread phenomenon. Five proteins that reportedly form bright fluorescent foci based on the ASKA[20] and the Venus[21] libraries were chosen and re-tagged with monomeric mGFPmut3. The *E. coli* cells were grown at 30 ˚C to exponential phase and imaged on agar pads with epi-fluorescence illumination. Bright fluorescent foci, resembling e.g. the ClpP-Venus foci, were not observed for any of the tested fusion proteins: (**a**) Hfq-mGFPmut3, (**b**) PepP-mGFPmut3, (**c**) IbpA-mGFPmut3, (**d**) MviM-mGFPmut3 and (**e**) FruK-mGFPmut3. The scale bars (white) are 1 µm.

We then further tested several other types of systems where individual studies have reported foci. For example, a SeqA-YFP fusion[141] forms artifactual foci in a *dam* deletion strain. SeqA controls the synchronicity of chromosomal replication in *E. coli*. The SeqA protein is a homotetramer that can form higher molecular-weight multimers *in vitro* in a concentration-dependent manner[142]. *In vivo*, SeqA binds to hemimethylated DNA and prevents inappropriate re-initiation at the origin. In a *dam* deletion strain, DNA is

permanently non-methylated and hence not recognized by SeqA. It was shown by immunofluorescence that SeqA is dispersed throughout the whole cell in such a dam deletion strain[45]. Since conjugation transfers only single-stranded DNA, foreign DNA introduced by conjugation from a *dam*+ strain into a *dam* deletion strain becomes permanently hemimethylated in the recipient cell and SeqA binds to this foreign DNA. Translational fusions between SeqA and a fluorescent protein, like YFP, then allow the visualization of the conjugation process by the formation of distinct foci in the recipient[141]. We found that a *dam* deletion strain containing the exact SeqA-YFP fusion (from reference [141]) occasionally produces foci in 10–18% of the cells even when no genetic transfer has occurred (**Figure 2.23a**–**c**). This observation is in conflict with the data reported in the Babic *et al.* paper[141] but does not invalidate their main conclusions. The YFP used for the SeqA fusion contains the following mutations (F46L S65G V68A S72A S175G T203Y) but lacks the monomeric A206K mutation[31].

We showed that a *dam* deletion strain containing a SeqA-mGFPmut3 fusion (Raul Fernandez-Lopez *et al*., unpublished) does not have this problem and forms no foci in the absence of genetic transfer (**Figure 2.23d**–**f**). As expected, both fusions, SeqA-YFP[141] and our SeqA-mGFPmut3, still form fluorescent foci in the recipient cells after conjugation (data not shown).

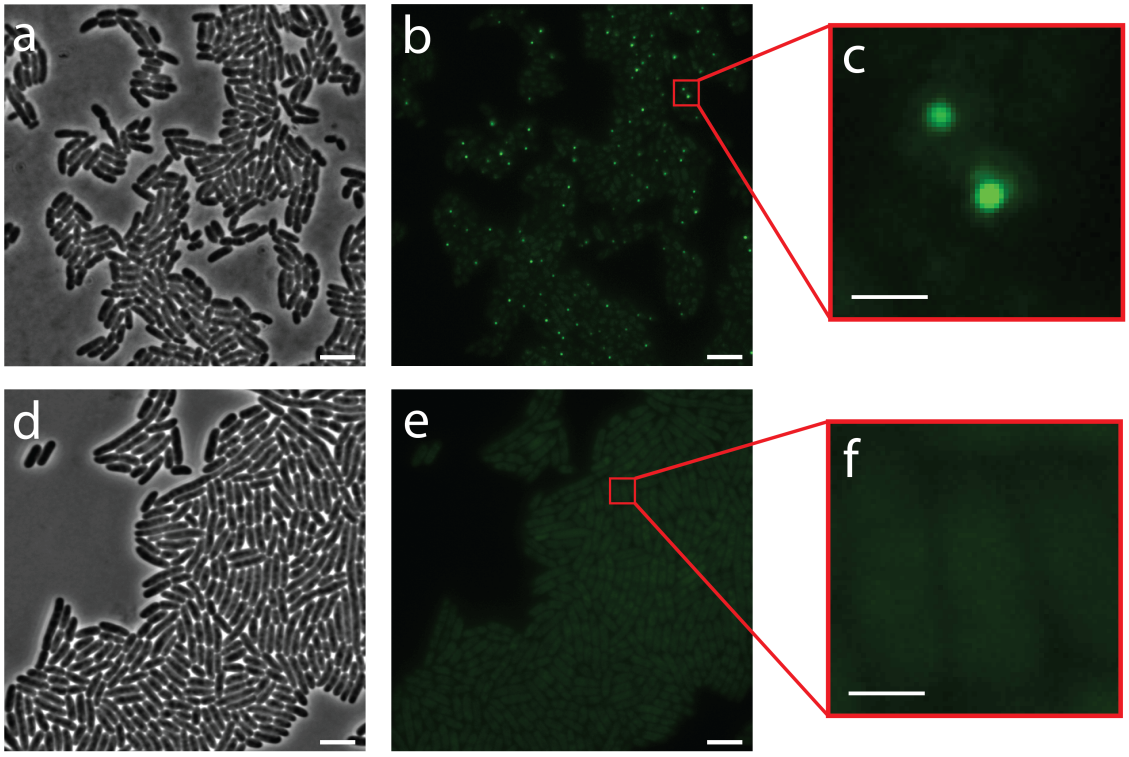


Figure 2.23. SeqA-YFP fusion forms foci even in the absence of Dam methylated DNA. (**a**) Phase-contrast image of *E. coli* cells expressing SeqA-YFP, (**b**) the corresponding YFP image and (**c**) close-up of two cells with bright fluorescent SeqA-YFP foci. A fusion between SeqA and the monomeric GFP variant, mGFPmut3, does not have this artifactual foci formation problem. (**d**) Phase-contrast image of cells expressing SeqA-mGFPmut3, (**e**) the corresponding GFP image and (**f**) a close-up of cells with a uniform GFP signal. Scale bars (white) correspond to 5 μm (subpanels a, b, d and e) or 1 μm (subpanels c and f).

Many groups also use repeats of DNA-binding sites and FP fusions to the corresponding DNA-binding protein to visualize the location of specific DNA segments (e.g. chromosomes, gene loci or plasmids). By bringing multiple FPs into close proximity, such systems could also be subject to avidity effects and FP multimerization due to high local concentrations. As observed for chromosomal loci[143], bacterial plasmids containing a *lacO* array with 240 LacI binding sites ($lacO_{240x}$ array) can be visualized in *E. coli* by co-expressing a LacI-FP fusion protein. Many copies of the LacI-FP protein bind to the

lacO array and create a diffraction-limited focus that is brighter than cellular autofluorescence (and background signal from cytoplasmic LacI-FP) and can be detected with normal epi-fluorescence microscopy[143].

We used *E. coli* cells harboring a pSC101 plasmid with a $lacO_{240x}$ array and a temperature-sensitive replication control system, which allowed us to reduce the plasmid copy number greatly by temporarily growing the cells at 42 ˚C. The cells therefore contained between zero and a few plasmids when they were imaged. I observed that the plasmids localize to the cell poles in a diffraction-limited focus for the LacI-EYFP fusion (**Figure 2.24a**). Polar localization of the pSC101 plasmids was also observed with the LacI-mYPet fusion (**Figure 2.24b**) but the apparent foci were sometimes larger (**Figure 2.24c**), and occasionally up to three diffraction-limited foci per cell pole could be observed (**Figure 2.24d**). Multiple discrete foci per pole were only observed with the fusion of LacI to the more monomeric fluorescent protein mYPet. This finding is consistent with the hypothesis that the LacI-EYFP fusion drives all plasmids into a single diffraction-limited focus because of fluorescent protein-mediated clustering, though in principle the differences between the LacI-EYFP and LacI-mYPet fusions could also be explained by a different sensitivity of the fluorescent proteins to the 42 ˚C step and potential folding or aggregation artifacts. mYPet is also not as monomeric as the mGFPs. Further work will be necessary to improve the LacI *lacO* system for imaging of plasmids in bacterial cells.

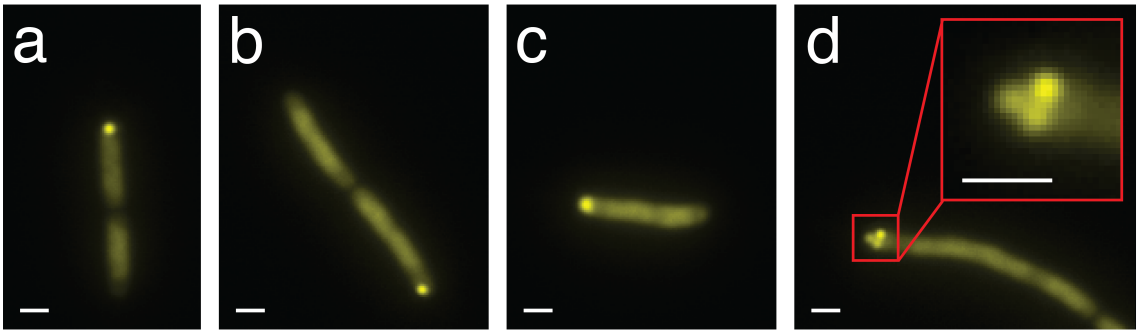


Figure 2.24. LacI-fluorescent protein fusions can aggravate clustering of pSC101 plasmids in *E. coli* cells. (**a**) The pSC101 plasmids localize to the cell poles in a diffraction-limited focus for the LacI-EYFP fusion. (**b**) Polar localization of the pSC101 plasmids was also observed with the LacI-mYPet fusion but (**c**) the foci appeared sometimes more smeared out and (**d**) occasionally up to three diffraction-limited foci per cell pole were observed. Multiple discrete foci per pole were only observed with the fusion of LacI to the monomeric fluorescent protein mYPet. Scale bars (white) correspond to 1 µm.

Single mRNA molecules can be detected in live *E. coli* cells with the MS2 system[32,144] but this requires that the mRNA of interest is tagged with an array of 96 binding sites for the MS2-GFP fusion protein. Tagged mRNAs usually co-localize into a single diffraction-limited focus per cell. I speculate that this is also because of avidity effects since the native MS2 proteins form a large oligomer that is composed of 180 MS2 molecules (i.e. the virion shell of bacteriophage MS2). Consistent with this hypothesis, preliminary experiments with an aggregation-deficient version of PP7[145] in combination with mEGFP does not show any sign of focus localization when imaging the *clpX* mRNA in live *E. coli* cells (see Chapter 3).

Taken together, our results with different systems strongly suggest that fluorescent protein-mediated clustering is a widespread phenomenon, although further tests, as presented for the Clp proteins, will be necessary to prove this unequivocally for each specific system.

2.3. Discussion

Fluorescent proteins are ubiquitously used in cell biology to study protein localization and have contributed tremendously to the understanding of sub-cellular organization. However, results obtained with fluorescent proteins in live cells are often blindly trusted and rarely verified with independent methods. Validation of protein localization measurements is a difficult task because of many technical complications and limited methodology. Here, I have introduced a new non-intrusive validation method for protein localization patterns in live cells. The method is based on the fact that spatial localization patterns affect the statistical differences between the two daughters after cell division. Measuring the post-division heterogeneity in some downstream process that depends on the protein of interest, with and without a reporter tag on the protein, thus allows for a direct equivalence test: if the downstream heterogeneity is different we know for a fact that the tag is intrusive.

I applied this assay to the bacterial Clp proteases, which were previously reported by many studies[20,21,23-30] to localize in diffraction-limited foci, which were argued to have important biological roles. My single-cell assay revealed that the focus formation of the oligomeric Clp proteases in *E. coli* is an artifact of the fluorescent proteins used. The Clp proteases are in fact uniformly distributed throughout the entire cell and thus, not localized in a focus. This finding is perfectly supported by the immunofluorescence in which co-localization with a foci-forming Venus fusion was used to test the antibody specificity. I also used SNAP and HALO tags, which are both unrelated to each other and unrelated to FPs, as well as HILO microscopy with a monomeric GFP that I found to be non-sticky. All independent methods, which were optimized with internal controls,

show that, surprisingly, fusing a spatially uniform FP to a spatially uniform native protein can cause severe foci formation.

I therefore used the ClpP system to evaluate twenty different protein tags and found that many tags, including supposedly monomeric fluorescent proteins, are prone to artifactual foci formation presumably because of a weak residual oligomerization tendency that are aggravated by local concentration effects (see below). I also showed that the artifactual Clp foci are large functional clusters and not mere conventional protein aggregates since the Clp foci are proteolytically active and highly fluorescent. Some tags, like monomeric GFP or the SNAP tag, do not produce foci and are less prone to clustering artifacts. Tagging the protease subunits, ClpA, ClpP and ClpX, with mGFPmut3 allowed visualization of the rapid intracellular movement of the Clp oligomers in real time (**Figure 2.17** and **Videos 2.4**–**2.7**).

Wild-type GFP forms dimers in a concentration dependent manner. This observation was already reported one year after the discovery of GFP as a tag[146] and the dissociation constant was later measured to be 100 μM[147]. The red fluorescent protein DsRed even forms an obligate tetramer and probably also higher-order oligomers[148]. Besides the oligomerization problem, DsRed also tends to cause protein aggregation[149,150]. Oligomerization of fluorescent proteins was reported to cause mislocalization artifacts in mammalian cells for fusion proteins targeted to the plasma membrane[31] and the endoplasmic reticulum[125,151] or fusions to the hexameric membrane protein connexin-43[82] and dimeric thymidine kinase[152]. Mislocalization artifacts due to fluorescent protein oligomerization were also evaluated in two recent small scale studies in mammalian cells[125,153] and discussed in a several reviews[76,154-157].

All GFP and DsRed derived fluorescent proteins are oligomers but some fluorescent proteins have been engineered into monomeric species by random mutagenesis often in

combination with a rational design strategy to break protein-protein interaction surfaces[154,156]. These new monomeric variants usually have the letter "m" (indicating that the protein is a monomer) preceding the name. GFP can be made monomeric with the A206K mutation, which introduces a positive charge in the hydrophobic dimer interface of GFP and increases the dissociation constant by almost three orders of magnitude[31] (**Figures 2.25a** and **2.25b**). The DsRed tetramer has two distinct dimer interfaces[158] (**Figure 2.25c**). Introducing 33 amino acid changes at the dimer interface and elsewhere in the proteins converted DsRed into the monomeric red fluorescent protein mRFP1[159]. Another 11 amino acid substitutions and the addition of new N- and C-termini resulted in the monomeric red fluorescent protein mCherry[70], which is brighter, faster maturing, more photo-stable and supposedly also more monomeric than mRFP1[70].

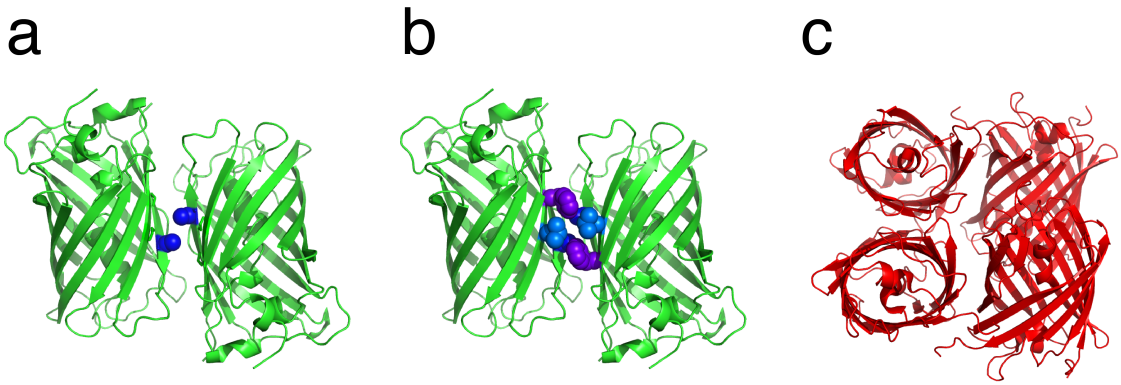


Figure 2.25. Many fluorescent proteins are weak oligomers. (**a**) GFP (green) forms a weak anti-parallel dimer[67]. Exchanging the hydrophobic amino acid alanine at position 206 (blue) of the dimer interface against a positively charged lysine residue reduces the dimerization tendency greatly[31]. (**b**) Mutations of two other hydrophobic amino acids, leucine 221 (light blue) and phenylalanine 223 (purple), also decreases dimer formation but the effect is less pronounced[31]. (**c**) DsRed is an obligate tetramer. Very extensive mutagenesis and protein reengineering resulted in the monomeric variants mRFP1[159] and mCherry[70]. The PDB entry for the DsRed structure is 1G7K.

The fluorescent protein survey with the Clp fusions showed that the presence of the A206K mutation in GFP or in a GFP-derived fluorescent protein, like Venus YFP,

abolished or greatly reduced foci formation (**Figure 2.13**). The A206K mutation seems to have a larger effect for the GFPs than for the YFPs, probably because these proteins have slightly different dimer interfaces with different residues making contacts[160].

Foci were still observed for the DsRed-derived fluorescent proteins, like mCherry and mCherry2, indicating that mCherry is less monomeric than commonly assumed.

The strongest foci formation was observed with the fluorescent protein superfolder GFP, which is also claimed to be a monomer[112]. Compared to wild-type GFP, superfolder GFP has a valine at position 206 and valine is a more hydrophobic amino acid than alanine. This increase in hydrophobicity probably explains the strong clustering tendency of superfolder GFP when fused to ClpP or ClpX. Mutating valine 206 to lysine (i.e. V206K mutation), led to monomeric superfolder GFP (msfGFP) variant and tagging ClpP with msfGFP did not result in foci formation. A single amino acid substitution thus resulted in a very striking effect. The ClpP-msfGFP fusion also performed well in the segregation assay (data not shown). The fluorescent protein sfGFP is ~1.6x brighter than EGFP (the same appears to be true for msfGFP), has similar photo-stability[68], matures fast (maturation half-time is 5.6 min, measured at 30 ˚C in yeast for normal sfGFP)[161] and is hence an excellent tag for protein fusions. A recent study also found that msfGFP is a non-intrusive tag for membrane-confined proteins and superior to EGFP, normal sfGFP and TagRFP, which all created mislocalization artifacts[125].

It is often not clear if the engineered monomeric fluorescent proteins are really true monomers or just less dimeric versions of the originals. In some cases, fluorescent proteins have been advertised as true monomers in the original publications but then later this assertion was proven wrong. For example, the red fluorescent protein TagRFP was initially advertised to be a monomer[162] but a later study demonstrated that TagRFP is indeed a weak dimer[163] and can create strong clustering artifacts[125]. The most commonly used super-resolution fluorescent protein mEos2[121] is also not a true

monomer, despite the “m” prefix, though a recent protein engineering effort developed mEos3.2[164], which is supposed to be truly monomeric. I found that TagRFP and mEos2 showed intermediate clustering and infrequent foci formation when fused to ClpP. This is in good agreement with the more recently reported weak oligomerization tendencies of these two fluorescent proteins.

Dissociation constants of fluorescent proteins are not always measured. When they are measured, the oligomerization state of the fluorescent protein is determined in a test tube in isolation where the fluorescent protein has no additional tag or fusion partner. This test condition is obviously quite different from the *in vivo* scenario where the fluorescent proteins diffuse through the crowded cytoplasm and, at least for protein localization measurements, are fused to other proteins. More importantly, since the fluorescent proteins are analyzed in isolation, any avidity issues that could originate from fusing an FP to a protein that oligomerizes or binds in large numbers to adjacent targets are not accounted for. Some reports evaluated the dimerization tendency of a fluorescent protein by constructing fusions to well-known model proteins, like tubulin, histones or actin, which have known localization in mammalian cells and are considered difficult to tag. For example, fluorescent protein fusions to tubulin often do not localize correctly if an oligomerization tendency is present[116]. Though these tests are mainly evaluated qualitatively (assessing whether localization is correct or not), the test fusions are usually also co-expressed in cells that contain the untagged version of the protein, which might act as a buffer and prevent weak oligomerization artifacts.

We believe the reason the Clp proteins form foci when tagged to FPs is that individual complexes collide, either due to simple diffusion in the cytoplasm or due to some natural tendency of binding to each other (**Figure 2.26**). An otherwise transient interaction is then stabilized by sticky fluorescent proteins, which in turn facilitates additional binding events between FPs. The phenomenon is thus cooperative where each FP interaction

helps create local concentration effects that promote further FP interactions. This explains why the FPs do not localize when expressed alone or when tagged to a monomeric protein, but create large foci when fused to proteins that naturally come in close proximity due to e.g. oligomerization or adjacent binding on nucleic acids or membranes.

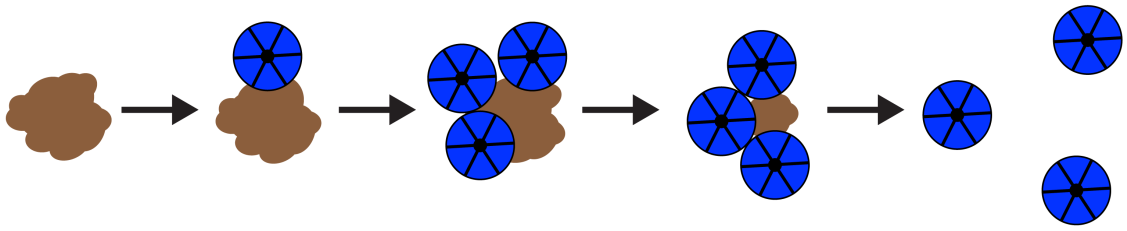


Figure 2.26. Spontaneous accumulation of protein aggregates could act as a nucleation seed for the Clp proteases to form foci. A few Clp proteases (blue) might bind to an intracellular protein aggregate (brown) and transiently co-localize. The proteases diffuse apart after degradation or disassembly of the aggregate but oligomeric fluorescent protein tags, even with very weak affinities, could interfere with the dissociation step and hence lead to the formation of bright fluorescent foci.

Fluorescent protein fusions that form foci are frequently observed in bacteria, as illustrated both by the large in-depth analyses of specific proteins and by the three genome-wide bacterial protein fusion libraries, which report that between 8–15% of the tested protein fusions localize in foci (**Table 2.1**). However, few of these foci were ever validated with another independent method. Here I showed that the same fluorescent proteins that were used to construct those libraries also produced artifactual foci when fused to ClpP or ClpX. My data in fact suggests that artifactual foci formation is a widespread problem that has been overlooked in the past. I re-analyzed the localization of a diverse set of *E. coli* proteins (Hfq, PepP, IbpA, MviM, FruK and SeqA), which are all oligomers (at least dimers) and were previously reported to from foci (**Table 2.2**). Retagging of these six proteins with the monomeric mGFPmut3 resulted in either no foci

or greatly reduced foci formation. This indicates that the previous reported foci were either artifacts or at least aggravated by the fluorescent proteins, although further work will be necessary to clarify this for each specific foci.

I also re-investigated the *lacO* array and PP7/MS2 system with more monomeric tags and obtained very different localization patterns than previous reports, which used more oligomeric tags. We have thus analyzed more than 10 systems in total, all chosen because the foci were established in multiple reports or in-depth studies, and in each case the patterns were caused by the FPs.

Foci formation of fluorescent protein fusions is less frequently observed in eukaryotic model systems like yeast or tissue culture cells. Just 3% (141 out of 4156 total) of the protein fusions of the GFP protein fusion library of *S. cerevisiae*[58] are classified as "punctate composite", which corresponds to protein fusions that form foci and this includes localization patterns with single and multiple foci. The frequency of foci observations is thus smaller than what was reported for the bacterial fluorescent protein fusion libraries despite the fact that the GFP that was used for the *S. cerevisiae* library also lacks the A206K mutation. Potential explanations for this discrepancy are that the GFP used might be quite monomeric despite the lack of the A206K mutation or the fact that yeast has less homo-oligomeric complexes than bacteria. For example, the bacterial 20S proteasome forms a barrel-shaped oligomer, reminiscent of the ClpXP protease, and is composed of two different subunits (α, β)[165] whereas the yeast 20S proteasome, which has a very similar structure, is composed of 14 different subunits (α_1–α_7, β_1–β_7)[166]. Only fluorescent protein fusions to homo-oligomeric proteins (but not hetero-oligomers) should result in high local concentrations and potential avidity effects, which could lead to foci formation of the respective fluorescent protein fusions. Bacteria are also more

densely packed with proteins and nucleic acids than eukaryotic cells. This crowding effect is well illustrated by the GFP diffusion constant, which is 7.7 $\mu m^2/s$ in *E. coli*[167], 27 $\mu m^2/s$ in mammalian cells[168] and 87 $\mu m^2/s$ in water[168]. A more crowded bacterial cytoplasm will favor association reactions and could at least partially explain the prevalence of foci in bacterial model systems.

It is hard to tell how many of the foci are real in cells, but a few foci-forming fluorescent protein fusions (e.g. [45,78]) have already been convincingly demonstrated by independent methods. However, once foci form, I expect them to be greatly exaggerated by the FP interactions. For these reasons I expect that virtually no foci can be quantitatively analyzed by FP methods, unless the most monomeric tags are used (ideally small peptide tags or single amino acid labels) and the results are quantitatively validated. On the other hand, I expect that virtually all foci are triggered by some initial spatial co-localization of proteins that drive larger foci formation. This applies to proteins that tend to cluster or are aggregation-sensitive, like chaperones, heat shock proteins[169], prions or amyloid proteins[170], but simply forming a multimeric complex is enough, as demonstrated by the Clp results.

Clustering artifacts may also affect studies of cell-to-cell heterogeneity. A recent theoretical analysis[171] showed that the increased partitioning error at cell division that is associated with clustering can create statistical patterns that perfectly mimic those associated with gene expression 'bursts'. FP clustering could then also confound studies of single-cell transcription and translation. On the other hand, the foci-forming Clp fusion constructs could be used as a tool in heterogeneity studies, to infer system properties by providing a means of modifying fluctuations without changing gene expression parameters, average abundances or even significantly the half-lives of the components.

The ClpP-sfGFP foci could for example be used to create cells that temporally lack (or have very low levels of) the ClpP protease after cell division and to analyze changes in substrate abundances or homeostatic control of the protease level.

The fluctuation-based assay described herein cannot be used for all cellular components since one cannot always measure the heterogeneity in an affected downstream process with the existing repertoire of single-cell reporters. However, using either FISH for detecting mRNAs or fluorescent protein reporters for proteins, the single-cell segregation assay can be used for any factor that directly or indirectly affects transcription, translation, RNA degradation or proteolysis. This includes many proteins that form foci, as well as mRNAs, for which specific mRNA localization has been reported with e.g. the MS2 system[32], a system that is prone to avidity effects (as discussed above). It may also be possible to measure other types of dependent processes, using light microscopy to measure division times or cell morphology, or using FRET biosensors to measure pH, metabolites, ATP levels, Ca^{2+} ions etc. Because the assay is based on a relative comparison with and without the upstream tag, while all other influences are unchanged, it is also insensitive to systematic measurement errors and can resolve small statistical differences. For example, if two different FPs produce different localization patterns (as expected from the fluorescent protein survey, **Figure 2.13**), testing which FP interferes less with the heterogeneity of a downstream process can suggest which reporter is more trustworthy.

2.4. Materials and Methods

Construction of *E. coli* strains

All *E. coli* strains used in this study are listed in **Table 2.3**.

Table 2.3

Strains	Description	Antibiotic marker	Reference
MC4100	Wild-type (wt) *E. coli* strain	-	Silhavy lab
DH5α	Standard cloning strain	-	Silver lab
CNP154	MC4100 λφ(*rpoS750*-LacZ) Δ*clpX*::Kan	Kan	Silhavy lab
DHL193	MC4100 pKD46	Amp	This study
DHL362	MC4100 pPM16	Amp	This study
DHL409	DH5α pDHL409	Amp	This study
DHL436	MC4100 *clpP*-Venus	-	This study
DHL440	MC4100 pDHL439	Amp	This study
DHL466	MC4100 *clpX*-Venus-T1 terminator	-	This study
DHL476	DHL466 pDHL439	Amp	This study
DHL477	DHL466 pDHL468	Amp	This study
DHL481	MC4100 pDHL468	Amp	This study
DHL524	MC4100 *clpX*-mCherry2-T1 terminator	-	This study
DHL526	MC4100 *clpX*-sfGFP-T1 terminator	-	This study
DHL534	DHL526 pDHL439	Amp	This study
DHL537	DHL436 pDHL439	Amp	This study
DHL564	MC4100 Δ*clpX*::Kan	Kan	This study
DHL620	MC4100 *clpX*-mGFPmut3	-	This study
DHL622	MC4100 *clpX*-GFP(–30)	-	This study
DHL624	MC4100 *clpX*-SNAP tag	-	This study
DHL626	MC4100 *clpX*-Dronpa	-	This study
DHL657	MC4100 *clpA*-mGFPmut3	-	This study
DHL659	MC4100 *clpA*-SNAP tag	-	This study
DHL661	MC4100 *clpP*-mGFPmut3	-	This study
DHL663	MC4100 *clpP*-SNAP tag	-	This study
DHL708	MC4100 Δ*clpPX*	-	This study
DHL722	DHL708 pDHL439	Amp	This study
DHL726	MC4100 *clpP*-Dronpa	-	This study
DHL747	DHL661 pDHL439	Amp	This study
DHL771	MC4100 *clpX*-rsFastLime	-	This study
DHL772	MC4100 *clpX*-PS-CFP2	-	This study
DHL778	MC4100 *clpP*-sfGFP	-	This study
DHL780	MC4100 *clpP*-TagRFP-T	-	This study
DHL785	MC4100 Δ*clpX*::P_{clpX}-mGFPmut3	-	This study

Table 2.3 (Continued).

DHL787	MC4100 *clpX*-mYPet	-	This study
DHL789	MC4100 *clpX*-TagRFP-T	-	This study
DHL805	DHL663 pDHL439	Amp	This study
DHL848	MC4100 *clpX*-mCherry-T1 terminator	-	This study
DHL922	MC4100 *hfq*-mGFPmut3	-	This study
DHL923	MC4100 *fruK*-mGFPmut3	-	This study
DHL925	MC4100 *mviM*-mGFPmut3	-	This study
DHL926	MC4100 *ibpA*-mGFPmut3	-	This study
DHL929	MC4100 *clpX*-mEos2	-	This study
DHL930	MC4100 *clpX*-mTagBFP	-	This study
DHL931	MC4100 *clpP*-mEos2	-	This study
DHL933	MC4100 *clpP*-Dendra2	-	This study
DHL944	MC4100 *pepP*-mGFPmut3	-	This study
DHL954	DHL726 pDHL439	Amp	This study
DHL955	DHL778 pDHL439	Amp	This study
DHL956	DHL933 pDHL439	Amp	This study
DHL986	DHL661 *att*Tn7::P_{clpPX}-*clpP*-mGFPmut3	-	This study
DHL987	DHL986 pDHL439	Amp	This study
DHL989	MC4100 *clpP*-mVenus	-	This study
DHL990	MC4100 *clpP*-mKate2	-	This study
DHL991	MC4100 *clpP*-mYPet	-	This study
DHL993	MC4100 *clpP*-mCherry	-	This study
DHL1000	MC4100 *clpP*-HaloTag	-	This study

The fluorescent protein fusions to *clpA*, *clpP* and *clpX* were constructed at the endogenous chromosomal loci with the lambda Red-mediated homologous recombination method[172]. Primers containing 50 nucleotide upstream or downstream homology to the integration site (*clpA*: DHL_P215_F and DHL_P216_R, *clpP*: DHL_P182_F and DHL_P183_R and *clpX*: DHL_P170_F and DHL_P171_R) were used to amplify the integration cassettes with the polymerase chain reaction (PCR) and AccuPrime Pfx (Invitrogen) or Phusion (Finnzymes) DNA polymerases. The PCR products were purified, digested with DpnI (NEB) and electroporated into *E. coli* strain DHL193 (MC4100 pKD46). Cells were resuspended in SOC media supplemented with 0.2% (w/v) arabinose, grown for 2 h at 30 ˚C (or 2 h at 30 ˚C followed by overnight incubation at room temperature) and spread on agar plates containing 30 μg/ml Kanamycin. Resistant colonies were isolated by restreaking and the presence of the

integration cassette was confirmed by colony PCR (http://openwetware.org/wiki/Knight:Colony_PCR). Two different primer sets, one for the upstream and one for the downstream integration scar, were used to confirm the site-specific integration. For both primer sets, one primer binds to the genome and the other one to the integration cassette (*clpA*_upstream: DHL_P313_F and NL73_R, *clpA*_downstream: DHL_P80_k3 and DHL_P314_R, *clpP*_upstream: DHL_P184_F and NL73_R, *clpP*_downstream: DHL_P80_k3 and DHL_P185_R, *clpX*_upstream: DHL_P173_F and NL73_R and *clpX*_upstream: DHL_P80_k3 and DHL_P174_R or primers DHL_P80_k3 and DHL_P556_R) and hence give no band for the wild-type strain. Plasmid pKD46 was eliminated from positive clones by growing the cells at 42 ˚C. Loss of the pKD46 plasmid was confirmed by testing the cells for sensitivity to ampicillin. The integrated fluorescent protein cassette was then P1 transduced into a fresh MC4100 strain using standard methods (http://openwetware.org/wiki/Sauer:P1vir_phage_transduction) followed by selection for kanamycin resistance on agar plates. The FRT-flanked Kan resistance marker cassette was then eliminated by transforming the cells with plasmid pCP20 and selection on agar plates with 100 µg/ml ampicillin at 30 ˚C. Elimination of the Kan marker, which leaves behind an 88 bp scar downstream of the stop codon of the fluorescent protein, was verified by PCR with gene-specific primers (*clpA*: DHL_P313_F and DHL_P314_R, *clpP*: DHL_P184_F and DHL_P185_R and *clpX*: DHL_P173_F and DHL_P174_R or primers DHL_P173_F and DHL_P556_R) and sequence-verified (Genewiz or in-house facility) with the same primer sets. Strains were also checked for sensitivity to kanamycin. Plasmid pCP20 was eliminated by growing the cells at 42 ˚C. Loss of the temperature-sensitive pCP20 plasmid was validated by testing for ampicillin sensitivity.

The following plasmids were used as PCR templates for constructing the chromosomal fluorescent protein fusions to *clpA*, *clpP* and *clpX*: pDHL414 for DHL436, pDHL146 for

DHL466, pDHL501 for DHL524, pDHL502 for DHL526, pDHL580 for DHL620, pDHL581 for DHL622, pDHL582 for DHL624, pDHL583 for DHL626, pDHL580 for DHL657, pDHL582 for DHL659, pDHL580 for DHL661, pDHL582 for DHL663, pDHL583 for DHL726, pDHL693 for DHL771, pDHL677 for DHL772, pDHL584 for DHL778, pDHL732 for DHL780, pDHL731 for DHL787, pDHL732 for DHL789, pDHL503 for DHL848, pDHL844 for DHL929, pDHL843 for DHL930, pDHL844 for DHL931, pDHL851 for DHL933, pDHL963 for DHL989, pBH27 for DHL990, pDHL731 for DHL991, and pDHL915 for DHL993.

Strain DHL564 was built by P1 transducing the Δ*clpX*::Kan allele from CNP154 into MC4100 and selection on agar plates containing 30 µg/ml kanamycin.

Strain DHL708 was built by deleting the *clpPX* operon with lambda-Red mediated homologous recombination (see above for details). The FRT-flanked Kan cassette was PCR amplified from pKD13 with primers DHL_P336_F and DHL_P337_R and transformed into DHL193. The Δ*clpPX* region was PCR amplified with primers DHL_P237_F and DHL_P174_R and the PCR product was sequenced.

Strain DHL785 was built by amplifying mGFPmut3 from pDHL580 with primers DHL_P375_F and DHL_P171_R and transforming the PCR product into strain DHL193 to replace the *clpX* coding region with mGFPmut3 by lambda-Red mediated homologous recombination. Expression of mGFPmut3 is hence under control of the endogenous P_{clpX} promoter in a Δ*clpX* background. The P_{clpX}-mGFPmut3 part was amplified with primers DHL_P184_F and DHL_P174_R, the PCR product was purified and the regions containing the integration scars where sequenced with primers DHL_P184_F and DHL_P174_R.

Plasmid pDHL580 was used as the PCR template for constructing the integration cassettes to tag *hfq*, *fruK*, *mviM*, *ibpA* and *pepP* with mGFPmut3, resulting in strains DHL922, DHL923, DHL925, DHL926 and DHL944. The strains were constructed

according to the lambda-Red protocol described above. Chromosomal integration was verified with two gene-specific primer sets, where one primer binds to the genome and the other primer to the integration cassette (*hfq*_upstream: DHL_P518_F and NL73_R, *hfq*_downstream: DHL_P80_F and DHL_P519_R, *fruK*_upstream: DHL_P520_F and NL73_R, *fruK*_downstream: DHL_P80_F and DHL_P521_R, *mviM*_upstream: DHL_P526_F and NL73_R, *mviM*_downstream: DHL_P80_F and DHL_P527_R, *ibpA*_upstream: DHL_P528_F and NL73_R, *ibpA*_downstream: DHL_P80_F and DHL_P529_R, *pepP*_upstream: DHL_P522_F and NL73_R and *pepP*_downstream: DHL_P80_F and DHL_P523_R). Verified integration cassettes were P1 transduced into MC4100 and the FRT-flanked Kan resistance cassette was removed. The integration sites were sequenced (Genewiz) with gene-specific primers (*hfq*: DHL_P518_F and DHL_P519_R, *fruK*: DHL_P520_F and DHL_P521_R, *mviM*: DHL_P526_F and DHL_P527_R, *ibpA*: DHL_P528_F and DHL_529_R and *pepP*: DHL_P522_F and DHL_P523_R).

Strain DHL986 was built by integrating a second copy of P_{clpPX}-*clpP*-mGFPmut3 in the *att*Tn7 site of strain DHL661 with a method previously described[173]. In short, strain DHL661 was transformed with plasmid pDHL970 and grown at 30 ˚C on Amp plates. Single colonies were isolated and the plasmid was eliminated by growth at 42 ˚C. Plasmid loss was confirmed by testing for ampicillin sensitivity. Chromosomal integration was verified by PCR with primers NL32_R and NL35_F. The fluorescence intensity of strain DHL986 is twice as high as the fluorescence intensity of DHL661 as expected for a strain that has two gene copies of *clpP*-mGFPmut3 (**Figure 2.14**).

Construction of plasmids

The plasmids and primers are listed in the Appendix in **Tables 2.4** and **2.5**, respectively.

Table 2.4

Plasmid	Description	Antibiotic marker	Reference
pBH27	pUC19-linker-mKate2-FRT Kan FRT	Amp, Kan	M. El Karoui (unpublished)
pCP20	Yeast Flp recombinase expression plasmid.	Amp	[172]
pDHL18	pUC19-linker-YFP-T1 terminator	Amp	This study
pDHL19	pUC19-FRT Kan FRT	Amp, Kan	This study
pDHL146	pUC19-linker-Venus-T1 terminator-FRT Kan FRT	Amp, Kan	This study
pDHL361	pUC19-mCherry2	Amp	This study
pDHL392	pUC19-linker-Venus	Amp	This study
pDHL409	pSC101-P_{LlacO1}-sfGFP-T1 terminator	Amp	This study
pDHL411	pSC101-$P_{A1/O4}$-Venus-ssrA(LAA) tag-T1 terminator	Amp	This study
pDHL414	pUC19-linker-Venus-FRT Kan FRT	Amp, Kan	This study
pDHL424	pSC101-$P_{A1/O4}$-mCherry-ssrA(LAA) tag-T1 terminator	Amp	This study
pDHL439	pSC101-$P_{A1/O4}$-mCherry-ssrA(LAA) tag-T1 terminator-*lacI*^{q}	Amp	This study
pDHL445	pUC19-linker-mCherry	Amp	This study
pDHL446	pUC19-linker-mCherry2-T1 terminator	Amp	This study
pDHL458	pUC19-linker-sfGFP-T1 terminator	Amp	This study
pDHL468	pSC101-$P_{A1/O4}$-mCherry-T1 terminator-*lacI*^{q}	Amp	This study
pDHL470	pUC19-linker-mCherry-T1 terminator	Amp	This study
pDHL501	pUC19-linker-mCherry2-T1 terminator-FRT Kan FRT	Amp, Kan	This study
pDHL502	pUC19-linker-sfGFP-T1 terminator-FRT Kan FRT	Amp, Kan	This study
pDHL503	pUC19-linker-mCherry-T1 terminator-FRT Kan FRT	Amp, Kan	This study
pDHL580	pUC19-linker-mGFPmut3-FRT Kan FRT	Amp, Kan	This study
pDHL581	pUC19-linker-GFP(–30)-FRT Kan FRT	Amp, Kan	This study
pDHL582	pUC19-linker-SNAP tag-FRT Kan FRT	Amp, Kan	This study
pDHL583	pUC19-linker-Dronpa tag-FRT Kan FRT	Amp, Kan	This study
pDHL584	pUC19-linker-sfGFP-FRT Kan FRT	Amp, Kan	This study
pDHL677	pUC19-linker-PS-CFP2-FRT Kan FRT	Amp, Kan	This study
pDHL693	pUC19-linker-rsFastLime-FRT Kan FRT	Amp, Kan	This study
pDHL731	pUC19-linker-mYPet-FRT Kan FRT	Amp, Kan	This study
pDHL732	pUC19-linker-TagRFP-T-FRT Kan FRT	Amp, Kan	This study
pDHL843	pUC19-linker-mTagBFP-FRT Kan FRT	Amp, Kan	This study
pDHL844	pUC19-linker-mEos2-FRT Kan FRT	Amp, Kan	This study
pDHL851	pUC19-linker-Dendra2-FRT Kan FRT	Amp, Kan	This study
pDHL915	pUC19-linker-mCherry-FRT Kan FRT	Amp, Kan	This study
pDHL963	pUC19-linker-mVenus-FRT Kan FRT	Amp, Kan	This study
pDHL970	pNDL1-P_{clpPX}-*clpP*-mGFPmut3	Amp	This study

Table 2.4 (Continued).

pKD13	Template plasmid for gene deletions. The Kan resistance gene is flanked by FRT sites.	Amp, Kan	[172]
pKD46	Lambda Red recombinase expression plasmid.	Amp	[172]
pNDL1	pGRG25 with *att*P sites for Gateway cloning in the multiple cloning site, pSC101 *ori ts*	Amp	N. Lord (unpublished)
pPM1	pSC101-*rep101*(ts)-P_{LtetO1}-Venus-T1 terminator	Amp	Lab collection
pPM14	pSC101-P_{LlacO1}-Venus-ssrA(LAA) tag-T1 terminator	Amp	Lab collection
pPM16	pSC101-P_{LlacO1}-Venus-T1 terminator	Amp	Lab collection
pPM88	pSC101-$P_{A1/O4}$-Venus-T1 terminator-*lacI*^{q}	Amp	Lab collection
pUC19	High copy number cloning vector	Amp	Invitrogen

Table 2.5

Primer	**Sequence**
DHL_P16_R	aagg-cccggg-ggcggatttgtcctactcaggag
DHL_P17_F	cctt-cccggg-gtgtaggctggagctgcttcgaag
DHL_P18_R	ggaagg-cctgcagg-ctgtcaaacatgagaattaattccggg
DHL_P21_F	cctt-gagctc-agcggtggcggtggc-agtaaaggagaagaacttttcactggagttg
DHL_P80_k3	gcccagtcatagccgaatagcc
DHL_P148_F	cg-ggatcc-atggttagtaaaggagaagaaaataacatgg
DHL_P149_R	ggaattc-aagctt-atgcggtaccagaacctttgtatagttc
DHL_P156_F	cg-ggatcc-atgtctaaaggtgaagaactgttcaccgg
DHL_P157_R	ggaattc-aagctt-atttgtagagctcatccatgccgtg
DHL_P158_F	agcggtggcggtggcagtaa
DHL_P159_R	attccggggatccgtcgacc
DHL_P165_R	aagg-cccggg-ttatttgtatagttcatccatgccatgtg
DHL_P170_F	aaccgttgctgatttatggcaagccggaagcgcaacaggcatctggtgaa-agcggtggcggtggcagtaa
DHL_P171_R	ggagataaaatccccccttttttggttaactaattgtatgggaatggttaa-attccggggatccgtcgacc
DHL_P173_F	ctgagtttattggtcgtctgccgg
DHL_P174_R	cgtcagtatatggggatgtttcccc
DHL_P175_R	ggaattc-gacgtc-ctgcggtaccagaacctttgtatagttc
DHL_P178_F	cctt-gagctc-agcggtggcggtggc-agtaaaggagaagaaaataacatggcaatc
DHL_P179_R	ggaatt-aagctt-atttgtatagttcatccatgccaccagtac
DHL_P180_F	cctt-gagctc-agcggtggcggtggc-agtaagggcgaggaggataacatggcc
DHL_P181_R	ggaatt-aagctt-acttgtacagctcgtccatgc
DHL_P182_F	ctgaagcggtggaatacggtctggtcgattcgattctgacccatcgtaat-agcggtggcggtggcagtaa
DHL_P183_R	agcgttgtgccgccctggataagtatagcggcacagttgcgcctctggca-attccggggatccgtcgacc

Table 2.5 (Continued).

DHL_P184_F	tttttgcctgccgaattcgc
DHL_P185_R	agcttgcgcacttcatgctgg
DHL_P186_F	cctt-gagctc-agcggtggcggtggc-agtaaaggtgaagaactgttcaccggtgttg
DHL_P189_R	ggaattc-aagctt-atttgtagagttcatccatgccgtg
DHL_P215_F	tgacttacggattccagagtgcacaaaagcacaaggcggaagcagcgcat-agcggtggcggtggcagtaa
DHL_P216_R	cgtaacctctttcgagattacggacttgaccaacctacctaacaatcaga-attccggggatccgtcgacc
DHL_P237_F	ccg-ctcgag-tttacgcagcataacgcgctaaattc
DHL_P252_R	aagg-cccggg-ttatttgtatagttcatccatgccatgtg
DHL_P253_F	cctt-gagctc-agcggtggcggtggc-agtaaaggtgaagagctgtttgacggtg
DHL_P254_R	aagg-cccggg-ttacttgtacagctcgtccattccatg
DHL_P255_F	cctt-gagctc-agcggtggcggtggc-agtaacgtgattaaaccagacatgaagatcaagc
DHL_P274_F	cctt-gagctc-agcggtggcggtggc-agtaacatggacaaagattgcgaaatgaaacg
DHL_P275_R	aagg-cccggg-ttaggagcctggcgcgcctatac
DHL_P276_R	aagg-cccggg-ttacttggcctgcctcggcag
DHL_P281_R	aagg-cccggg-ttatttgtagagttcatccatgccgtg
DHL_P297_F	cctt-gagctc-agcggtggcggtggc-agtaaaggtgaagaattattcactggtgttgtc
DHL_P298_R	ccagcctacacccgggttatttgtac
DHL_P306_F	cctt-gagctc-agcggtggcggtggc-agtaaagtgtctaagggcgaagagctgattaag
DHL_P307_R	aagg-cccggg-ttaattaagtttgtgccccagtttgctagg
DHL_P313_F	gccaggaagcgcgtaactgg
DHL_P314_R	cagtgaacatggtgggcggg
DHL_P321_F	cctt-gagctc-agcggtggcggtggc-agtaaaagtgcgattaagccagacatgaagatc
DHL_P322_R	aagg-cccggg-ttatcgtctggcattgtcaggcaatc
DHL_P336_F	ttacaatcggtacagcaggttttttcaattttatccaggagacggaaatg-attccggggatccgtcgacc
DHL_P337_R	tggttaactaattgtatgggaatggttaattattcaccagatgcctgttg-tgtaggctggagctgcttcg
DHL_P346_F	ggc-aacatggctctgtcgcttgaagg
DHL_P347_R	atcacccttcagcactccatcacg
DHL_P375_F	ttgcgtcgtcgtgtgcggcacaaagaacaaagaagaggttttgacccatg-agtaaaggagaagaacttttcactggagttg
DHL_P440_F	cctt-gagctc-agcggtggcggtggc-agtaagggcgccgagctgttcac
DHL_P441_R	aagg-cccggg-ttacttgtacagctcatccatgccg
DHL_P449_F	cctt-gagctc-agcggtggcggtggc-agtaacatgagcgagctgattaaggagaacatg
DHL_P450_R	aagg-cccggg-ttaattaagcttgtgccccagtttgc
DHL_P453_F	cctt-gagctc-agcggtggcggtggc-agtaacatgaacacaccgggaattaacc
DHL_P454_R	aagg-cccggg-ttaccacacctggctgggcagg
DHL_P472_R	aagg-cccggg-ttatttgtatagttcatccatgccaccagtac
DHL_P496_F	gtagcagcgcgcagaatacttccgcgcaacaggacagcgaagaaaccgaa-agcggtggcggtggcagtaa
DHL_P497_R	cggggaacgcaggatcgctggctccccgtgtaaaaaaacagcccgaaacc-attccggggatccgtcgacc
DHL_P498_F	gtccgcagttggccgcaatgatggcgcgcgtcgacttacaacctttaac-agcggtggcggtggcagtaa
DHL_P499_R	cgagattagcgtcaataatcagcagcgttttcattatgcctctcctgctg-attccggggatccgtcgacc

Table 2.5 (Continued).

DHL_P500_F	tgaaaaagccggaagaaatcgaagcgttgatggttgctgcgagaaagcaa-agcggtggcggtggcagtaa
DHL_P501_R	ggaaatagccagcgccagcgtcgcgcccgccatgccgccaccgacgatgattacgctcat-attccggggatccgtcgacc
DHL_P504_F	tgctggcgcaacgtatcgttgacaagatctggcgcgatgcgatgagtgaa-agcggtggcggtggcagtaa
DHL_P505_R	aatagtctacctggattatggtgaattgctaccgccagatgttacagggt-attccggggatccgtcgacc
DHL_P506_F	tcgaacgcgtgattccggaagcgaaaaaaccgcgccgtatcgaaatcaac-agcggtggcggtggcagtaa
DHL_P507_R	cctgacggcgagcatggagatgtcaggccgcgccaggcggccttagggaa-attccggggatccgtcgacc
DHL_P518_F	aaggaaaagagagaatggctaagggg
DHL_P519_R	cttccgcaatttcaactgctttacc
DHL_P520_F	gaatggatcgccaaaccaccg
DHL_P521_R	tcaggaacagctcagggtgcg
DHL_P522_F	gactggatgtccatgacgtggg
DHL_P523_R	ctgacatgcacggtggtgatgg
DHL_P526_F	gtcagcgtgaaaccgtgcagg
DHL_P527_R	cgtaacaagttaggaagtttaaaagcgacg
DHL_P528_F	gcattgctatcgctgtggctgg
DHL_P529_R	ttttctcaatgttgtacggcggg
DHL_P550_F	gtcacgacgttgtaaaacgacggccagtgaattcgagctcagcggtggcggtggcagtaa
DHL_P551_R	tctcttttcgttgggatctttcgaaagtttagattgataggacaggtaatggttgtctgg
DHL_P552_F	accagacaaccattacctgtcctatcaatctaaactttcgaaagatcccaacgaaaagag
DHL_P553_R	acttcgaagcagctccagcctacaccccgggttatttgtatagttcatccatgccatgtg
DHL_P556_R	atgacacgactgtgcttcacgc
DHL_P557_F	gggg-acaagtttgtacaaaaaagcaggctct-cgcagcataacgcgctaaattcg
DHL_P558_R	gggg-accactttgtacaagaaagctgggtc-attccggggatccgtcgacc
M13_R	caggaaacagctatgacc
NL15_F	ctggttggtcgacactagtattaccc
NL32_R	gatgacggtttgtcacatgga
NL35_F	cccctatagtgagtcgtattacatgg
NL73_R	ggattcatcgactgtggccg

All primers and DNA oligonucleotides were purchased from Integrated DNA Technologies, Inc. (IDT). The Plasmid constructions were verified by analytical restriction digests. All cloning steps that involved PCR amplification were validated by DNA sequencing. Vent (NEB) or Phusion (Finnzymes) polymerases were used for standard cloning. All restriction enzymes were purchased from NEB and used according to the manufacturer's instructions.

Two types of plasmids were used for constructing the chromosomal fluorescent protein fusions to *clpA*, *clpP* and *clpX*. Both plasmids were derived from the same ancestor (pDHL19) and are identical except that one carries the linker-fluorescent protein alone and the other one carries the linker-fluorescent protein followed by the T1 transcriptional terminator. All fluorescent protein-tagging vectors share the same primer binding sites for amplifying the fluorescent protein-FRT Kan FRT integration cassettes. The forward primer binding site (see DHL_P158_F for sequence) corresponds to the linker region and the reverse primer binding site (see DHL_P159_R for sequence) is downstream of the FRT-flanked Kan marker.

pDHL19 was built by amplifying the FRT-flanked kanamycin resistance marker cassette (FRT Kan FRT) from pKD13[172] with primers DHL_P17_F and DHL_P18_R, the PCR product was digested with XmaI and SbfI and ligated into pUC19, which was cut with the same restriction enzymes.

pDHL146 was built in two steps. Venus was amplified from pPM1 with primers DHL_P21_F and DHL_P16_R, the PCR product was digested with SacI and XmaI and ligated into pUC19, which was cut with the same enzymes. The resulting plasmid, named pDHL18, was digested with SacI and XmaI and the linker-Venus-T1 terminator fragment was purified and ligated into SacI/XmaI-digested pDHL19. The Venus[114] variant used here contains two additional mutations (V68L, Q69M) from Citrine[174,175]. These two mutations are known to be beneficial to Citrine.

pDHL409 was built by amplifying superfolder GFP (sfGFP)[112] from pHC467 (courtesy of Prof. T. Bernhardt, Harvard Medical School) with primers DHL_P156_F and DHL_P157_R, digested with BamHI and HindIII and ligated into pPM16, which was also digested with the same enzymes. This puts expression of sfGFP under the control of the P_{LlacO1} promoter[176], which is constitutively 'ON' in MC4100 (since this strains lacks the *lacI* gene). pPM16 harbors a pSC101 origin and the *bla* gene for Amp resistance.

pDHL414 was built in two steps. First, Venus[114] was amplified from pPM1 with primers DHL_P21_F and DHL_P165_R, digested with SacI and XmaI and subcloned into pUC19 cut with the same enzymes. The resulting plasmid, pDHL392, was digested with SacI and XmaI and the SacI-linker-Venus-XmaI fragment was purified and ligated into pDHL19, which was also cut with SacI and XmaI.

pDHL439 was built in three steps. First, pPM88 was digested with XhoI and EcoRI to cut out the $P_{A1/O4}$ promoter[177]. The DNA fragment was purified and ligated into XhoI/EcoRI-cut pPM14, resulting in plasmid pDHL411. Second, mCherry[70] without the stop codon was amplified from BioBrick Z0075[178] with primers DHL_P148_F and DHL_P175_R, digested with BamHI and AatII and ligated into pDHL411, which was digested with the same enzymes. The resulting plasmid, named pDHL424, was digested with XhoI and XbaI, the fragment corresponding to $P_{A1/O4}$-mCherry-ssrA(LAA) was purified and ligated into pPM88, which was also digested with XhoI and XbaI. pPM88 has a pSC101 origin, the *bla* gene and *lacI*^{q}. The ssrA(LAA) tag is the native *E. coli* ssrA tag with the following amino acid sequence: AANDENYALAA[90,91].

pDHL468 was derived from pDHL439 by exchanging the mCherry-ssrA(LAA) fragment for mCherry. mCherry was PCR amplified from pDHL439 with primers DHL_P148_F and DHL_P149_R, digested with BamHI and HindIII and ligated into pDHL439, which was digested with the same enzymes.

pDHL501 was built in two steps. Primers DHL_P180_F and DHL_P181_R were used to amplify mCherrry2 from pDHL361. mCherry2 was synthesized by GenScript (DNA sequence courtesy of Dr. N. Shaner). The PCR product was digested with SacI and HindIII. Next, pDHL18 was digested with HindIII and the HindIII-T1 terminator-HindIII part was purified. The SacI-linker-mCherry-HindIII and the HindIII-T1 terminator-HindIII fragments were subcloned into SacI/HindIII-digested pUC19 resulting in plasmid

pDHL446. The linker-mCherry2-T1 terminator fragment was then cut out of pDHL446 with SacI and XmaI and ligated into SacI/XmaI-cut pDHL19.

pDHL502 was built in two steps. First, primers DHL_P186_F and DHL_P189_R were used to amplify sfGFP[112] from pDHL409. Primer DHL_P189_R introduces a silent mutation into the C-terminus of sfGFP to eliminate an existing SacI site. The PCR product and pDHL18 were digested with SacI and HindIII. The digested PCR product, the HindIII-T1 terminator-HindIII fragment and the pDHL18 backbone were purified and combined in a triple ligation and the resulting plasmid was named pDHL458. The linker-sfGFP-T1 terminator fragment was then cut out from pDHL458 with SacI and XmaI and ligated into pDHL19, also cut with SacI and XmaI.

pDHL503 was built in two steps. Primers DHL_P178_F and DHL_P179_R were used to amplify mCherry[70] from BioBrick Z0075[178], the PCR product was digested with SacI and HindIII and subcloned into SacI/HindIII-digested pUC19 giving plasmid pDHL445. The mCherry in Z0075 already contained the D8N mutation. Next, pDHL18 was digested with HindIII and the HindIII-T1 terminator-HindIII part was purified. Vector pDHL445 was cut with SacI and HindIII, the linker-mCherry insert was purified and subcloned together with the T1 terminator part into SacI/HindIII-digested pDHL18. The resulting plasmid was named pDHL470. The linker-mCherry-T1 terminator fragment was released from pDHL470 by digestion with SacI and XmaI and ligated into SacI/XmaI-digested pDHL19.

pDHL580 was built in a single step. mGFPmut3 was amplified from pDH78 (D. Huh, Y. Jung and J. Paulsson, manuscript in preparation) with primers DHL_P21_F and DHL_P252_R, digested with SacI and XmaI and ligated into pDHL19, which was also digested with SacI and XmaI. mGFPmut3 corresponds to GFPmut3[179] with the A206K[31] mutation.

pDHL581 was built by PCR amplifying GFP(–30) from pET-GFP-NEG30[124] with primers DHL_P253_F and DHL_P254_R, digested with SacI and XmaI and ligated into pDHL19, which was digested with the same enzymes.

pDHL582 was built by amplifying the SNAP tag[108] from the pSNAP-tag(T7) plasmid (NEB, cat.# N9174S) with primers DHL_P274_F and DHL_P275_R. The PCR product was digested with SacI and XmaI and ligated into SacI/XmaI-digested pDHL19.

pDHL583 was built by amplifying Dronpa[127] from pcDNA3-Dronpa (courtesy of Dr. H. Zhong, Janelia Farm) with primers DHL_P255_F and DHL_P276_R, digested with SacI and XmaI and ligated into pDHL19, which was also digested with SacI and XmaI.

pDHL584 was built by amplifying sfGFP[112] from pDHL502 with primers DHL_P186_F and DHL_P281_R. Primer DHL_P281_R introduces a silent mutation into the C-terminus of sfGFP to eliminate an existing SacI site. The PCR product was digested with SacI and XmaI and ligated into pDHL19, which was digested with the same enzymes.

pDHL677 was built by amplifying PS-CFP2[122] from pPS-CFP2-N (Evrogen, cat.# FP802) with primers DHL_P440_F and DHL_P441_R, digested with SacI and XmaI and ligated into SacI/XmaI-digested pDHL19.

pDHL693 was built in two steps. First, primers DHL_P346_F and DHL_P347_R were used to introduce the V157G mutation in plasmid pET15b-Dronpa-AP tag (lab collection) with 'Round-the-horn site-directed mutagenesis (http://openwetware.org/wiki/'Round-the-horn_site-directed_mutagenesis) and Phusion (Finnzymes) DNA polymerase. The resulting vector was used as the PCR template to amplify rsFastLime[123] (i.e., Dronpa V157G) with primers DHL_P255_F and DHL_P276_R. The PCR product was then digested with SacI and XmaI and ligated into pDHL19, which was also cut with SacI and XmaI.

pDHL731 was built by amplifying mYPet from pROD50[118] (i.e., pUC-11aa-mYPet-kan) with primers DHL_P297_F and DHL_P298_R, digested with SacI and XmaI and ligated

into pDHL19, which was cut with the same enzymes. mYPet corresponds to YPet[180] with the A206K[31] mutation.

pDHL732 was built by amplifying TagRFP-T[116] from pmTagRFP-T-Tubulin-6 (courtesy of Prof. M. Davidson, Florida State University) with primers DHL_P306_F and DHL_P307_R, digested with SacI and XmaI and ligated into pDHL19, which was also digested with SacI and XmaI.

pDHL843 was built by PCR amplifying mTagBFP[119] (courtesy of Prof. M. Springer, Harvard Medical School) with primers DHL_P449_F and DHL_P450_R. The PCR product was digested with SacI and XmaI and ligated into pDHL19, which was also cut with SacI and XmaI.

pDHL844 was built by amplifying mEos2[121] from pRsetA-mEos2[121] with primers DHL_P321_F and DHL_P322_R, digested with SacI and XmaI and ligated into SacI/XmaI-digested pDHL19.

pDHL851 was built by PCR amplifying Dendra2[128] from plasmid pDendra2-Tubulin (courtesy of Sophie Dumont, Harvard Medical School) with primers DHL_P453_F and DHL_P454_R. The insert was digested with SacI and XmaI and then ligated into SacI/XmaI-digested pDHL19.

pDHL915 was built by amplifying mCherry[70] from pDHL503 with primers DHL_P178_F and DHL_P472_R. The PCR product was digested with SacI and XmaI and ligated into pDHL19, which was cut with the same restriction enzymes.

pDHL963 was built in two steps. First, Venus[114] part 1 (aa# 2–205) was amplified from pDHL414 with primers DHL_P550_F and DHL_P551_R and Venus part 2 (aa# 207–238) was amplified from pDHL414 with primers DHL_P552_F and DHL_P553_R. Primers DHL_P551_R and DHL_P552_F have the A206K[31] in the overhangs. The PCR products were purified and inserted into the SacI/XmaI-digested pDHL19 vector with isothermal assembly[181].

pDHL970 was built by PCR amplifying the region corresponding to P_{clpPX}-*clpP*-mGFPmut3 from genomic DNA of strain DHL661 with primers DHL_P557_F and DHL_P558_R. The PCR product was cloned into pNDL1 with Gateway cloning (Invitrogen) and sequence-verified with primers M13_R, NL15_F and DHL_P184_F.

Fluorescent proteins, SNAP tag and HaloTag

The fluorescent proteins, SNAP tag and HaloTag used in this study, including their amino acid sequences, are listed in the Appendix in **Table 2.6**.

Table 2.6

Name	Additional mutations	Amino acid sequence	Reference
mTagBFP	ΔE3	(SGGGGSNM)-SELIKENMHMKLYMEGTVDNHHFKCTSEGEGKPYEGTQTMRIKVVEGGPLPFAFDILATSFLYGSKTFINHTQGIPDFFKQSFPEGFTWERVTTYEDGGVLTATQDTSLQDGCLIYNVKIRGVNFTSNGPVMQKKTLGWEAFTETLYPADGGLEGRNDMALKLVGGSHLIANIKTTYRSKKPAKNLKMPGVYYVDYRLERIKEANNETYVEQHEVAVARYCDLPSKLGHKLN	[119]
PS-CFP2	-	(SGGGG)-SKGAELFTGIVPILIELNGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVATLSYGVQCFSRYPDHMKQHDFFKSAMPEGYIQERTIFFEDDGNYKTRAEVKFEGDTLVSRIELTGTDFKEDGNILGNKMEYNYNATNVYIVADKARNGIKVNFKVRHNIKDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMIYLEFVTAAAITHGMDELYK	[122]
Superfolder GFP (sfGFP)	-	(SGGGG)-SKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDELYK	[112]

Table 2.6 (Continued).

msfGFP	V206K	sfGFP with V206K	This study
mGFPmut3	A206K	(SGGGG)- SKGEELFTGVVPILVELDGDVNGHKFSVSGE GEGDATYGKLTLKFICTTGKLPVPWPTLVTT FGYGVQCFARYPDHMKQHDFFKSAMPEGY VQERTIFFKDDGNYKTRAEVKFEGDTLVNRI ELKGIDFKEDGNILGHKLEYNYNSHNVYIMA DKQKNGIKVNFKIRHNIEDGSVQLADHYQQN TPIGDGPVLLPDNHYLSTQSKLSKDPNEKRD HMVLLEFVTAAGITHGMDELYK	This study
GFP(–30)	-	(SGGGG)- SKGEELFDGVVPILVELDGDVNGHEFSVRG EGEGDATEGELTLKFICTTGELPVPWPTLVT TLTYGVQCFSDYPDHMDQHDFFKSAMPEG YVQERTISFKDDGTYKTRAEVKFEGDTLVNR IELKGIDFKEDGNILGHKLEYNFNSHDVYITA DKQENGIKAEFEIRHNVEDGSVQLADHYQQ NTPIGDGPVLLPDDHYLSTESALSKDPNEDR DHMVLLEFVTAAGIDHGMDELYK	[124]
Dronpa	-	(SGGGGSN)- VIKPDMKIKLRMEGAVNGHPFAIEGVGLGKP FEGKQSMDLKVKEGGPLPFAYDILTTVFCYG NRVFAKYPENIVDYFKQSFPEGYSWERSMN YEDGGICNATNDITLDGDCYIYEIRFDGVNFP ANGPVMQKRTVKWEPSTEKLYVRDGVLKG DVNMALSLEGGGHYRCDFKTTYKAKKVVQL PDYHFVDHHIEIKSHDKDYSNVNLHEHAEAH SELPRQAK	[127]
rsFastLime	-	Dronpa with V157G	[123]
mEos2	-	(SGGGGSK)- SAIKPDMKIKLRMEGNVNGHHFVIDGDGTGK PFEGKQSMDLEVKEGGPLPFAFDILTTAFHY GNRVFAKYPDNIQDYFKQSFPKGYSWERSL TFEDGGICIARNDITMEGDTFYNKVRFYGTN FPANGPVMQKKTLKWEPSTEKMYVRDGVLT GDIHMALLLEGNAHYRCDFRTTYKAKEKGV KLPGYHFVDHCIEILSHDKDYNKVKLYEHAV AHSGLPDNARR	[121]
Dendra2	-	(SGGGGSNMNTPGI)- NLIKEDMRVKVHMEGNVNGHAFVIEGEGKG KPYEGTQTANLTVKEGAPLPFSYDILTTAVH YGNRVFTKYPEDIPDYFKQSFPEGYSWERT MTFEDKGICTIRSDISLEGDCFFQNVRFKGT NFPPNGPVMQKKTLKWEPSTEKLHVRDGLL VGNINMALLLEGGGHYLCDFKTTYKAKKVVQ LPDAHFVDHRIEILGNDSDYNKVKLYEHAVA RYSPLPSQVW	[128]

Table 2.6 (Continued).

Venus	V68L Q69M	(SGGGG)- SKGEELFTGVVPILVELDGDVNGHKFSVSGE GEGDATYGKLTLKLICTTGKLPVPWPTLVTT LGYGLMCFARYPDHMKRHDFFKSAMPEGY VQERTIFFKDDGNYKTRAEVKFEGDTLVNRI ELKGIDFKEDGNILGHKLEYNYNSHNVYITAD KQKNGIKANFKIRHNIEDGGVQLADHYQQNT PIGDGPVLLPDNHYLSYQSALSKDPNEKRDH MVLLEFVTAAGITHGMDELYK	[114]
mVenus	A206K	Venus with A206K	[117]
mYPet	A206K	(SGGGGSK)- GEELFTGVVPILVELDGDVNGHKFSVSGEGE GDATYGKLTLKLLCTTGKLPVPWPTLVTTLG YGVQCFARYPDHMKQHDFFKSAMPEGYVQ ERTIFFKDDGNYKTRAEVKFEGDTLVNRIELK GIDFKEDGNILGHKLEYNYNSHNVYITADKQ KNGIKANFKIRHNIEDGGVQLADHYQQNTPI GDGPVLLPDNHYLSYQSKLFKDPNEKRDHM VLLEFLTAAGITEGMNELYK	[118,180]
TagRFP-T	-	(SGGGGSK)- VSKGEELIKENMHMKLYMEGTVNNHHFKCT SEGEGKPYEGTQTMRIKVVEGGPLPFAFDIL ATSFMYGSRTFINHTQGIPDFFKQSFPEGFT WERVTTYEDGGVLTATQDTSLQDGCLIYNV KIRGVNFPSNGPVMQKKTLGWEANTEMLYP ADGGLEGRTDMALKLVGGGHLICNFKTTYR SKKPAKNLKMPGVYYVDHRLERIKEADKETY VEQHEVAVARYCDLPSKLGHKLN	[116]
mKate2	-	(SGGGGSK)- VSELIKENMHMKLYMEGTVNNHHFKCTSEG EGKPYEGTQTMRIKAVEGGPLPFAFDILATS FMYGSKTFINHTQGIPDFFKQSFPEGFTWER VTTYEDGGVLTATQDTSLQDGCLIYNVKIRG VNFPSNGPVMQKKTLGWEASTETLYPADG GLEGRADMALKLVGGGHLICNLKTTYRSKKP AKNLKMPGVYYVDRRLERIKEADKETYVEQ HEVAVARYCDLPSKLGHR	[115]
mCherry	D8N	(SGGGG)- SKGEENNMAIIKEFMRFKVHMEGSVNGHEF EIEGEGEGRPYEGTQTAKLKVTKGGPLPFA WDILSPQFMYGSKAYVKHPADIPDYLKLSFP EGFKWERVMNFEDGGVVTVTQDSSLQDGE FIYKVKLRGTNFPSDGPVMQKKTMGWEASS ERMYPEDGALKGEIKQRLKLKDGGHYDAEV KTTYKAKKPVQLPGAYNVNIKLDITSHNEDYT IVEQYERAEGRHSTGGMDELYK	[70]
mCherry2	-	mCherry (without D8N) with K97N K143C K74R S152T N201D T207L	N. Shaner (unpublished)

Table 2.6 (Continued).

SNAP tag (SNAP26b)	-	(SGGGGSN)-MDKDCEMKRTTLDSPLGKLELSGCEQGLHEIKLLGKGTSAADAVEVPAPAAVLGGPEPLMQATAWLNAYFHQPEAIEEFPVPALHHPVFQQESFTRQVLWKLLKVVKFGEVISYQQLAALAGNPAATAAVKTALSGNPVPILIPCHRVVSSSGAVGGYEGGLAVKEWLLAHEGHRLGKPGLGPAGIGAPGS	[108]
SNAP tag (SNAPf)	-	(SGGGGSN)-MDKDCEMKRTTLDSPLGKLELSGCEQGLHRIIFLGKGTSAADAVEVPAPAAVLGGPEPLMQATAWLNAYFHQPEAIEEFPVPALHHPVFQQESFTRQVLWKLLKVVKFGEVISYSHLAALAGNPAATAAVKTALSGNPVPILIPCHRVVQGDLDVGGYEGGLAVKEWLLAHEGHRLGKPGLG	[182]
HaloTag (HT7)	-	(SGGGGSND)-GSEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLHGNPTSSYVWRNIIPHVAPTHRCIAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLEEVVLVIHDWGSALGFHWAKRNPERVKGIAFMEFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQNVFIEGTLPMGVVRPLTEVEMDHYREPFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDWLHQSPVPKLLFWGTPGVLIPPAEAARLAKSLPNCKAVDIGPGLNLLQEDNPDLIGSEIARWLSTLEISG	[109]

Note: The amino acid sequence corresponding to the linker is surrounded by parentheses. For all fluorescent proteins, the linker replaces the first methionine, which is encoded by the start codon.

Western blot analysis

Overnight cultures of the respective *E. coli* strains were diluted 1:1,000 in LB medium and grown with shaking (220 rpm) at 37 ˚C until the cultures reached OD_{600} = 0.2–0.3. Then, 1,600 µl of cell suspension was pelleted (15,000 g, 1 min, room temperature), normalized by OD_{600} by resuspending the cell pellet in *x* µl (where *x* is *x* = 160 x OD_{600}) of 1x SDS loading buffer (80 mM Tris-HCl pH 6.8, 2% (w/v) SDS, 10% (v/v) glycerol, 5% (w/v) ß-mercaptoethanol, 0.01% (w/v) bromphenol blue) and boiled at 95 ˚C for 5 min. 10 µl cell lysate per lane and 5 µl molecular weight marker (SeeBlue Plus2 Pre-Stained Standard, Invitrogen, cat# LC5925) were loaded on a 10% SDS page (Tris-glycine) and

the proteins were separated. Next, the proteins were transferred to a 0.2-µm nitrocellulose membrane (Sigma-Aldrich, cat# N7892-5EA) by electroblotting (100 V, 1 h, 4 ˚C). After the transfer, the membrane was incubated in blocking buffer (1x TBS, 0.1% (v/v) Tween-20, 4% (w/v) milk powder) for 1 h at room temperature. Then, the membrane was incubated with a polyclonal rabbit anti-GFP antibody (courtesy of Prof. P. Silver, Harvard Medical School) for the Clp-fluorescent protein fusions or polyclonal anti-SNAP tag antibody (NEB, cat# P9310S) for the Clp-SNAP tag fusions, diluted 1:1,000 in antibody dilution buffer (1x TBS, 2% (w/v) BSA, 0.1% (v/v) Tween-20, 0.05% (w/v) NaN_3), for overnight at 4 ˚C. The membrane was washed 4 times for 20 min in TBST (1x TBS, 0.1% (v/v) Tween-20) with gentle shaking. Next, the membrane was incubated with anti-rabbit HRP-linked secondary antibody (GE Healthcare, NA934V), diluted 1:5,000 in blocking buffer, for 1 h at room temperature. Lastly, the membrane was washed 4 times for 5 min with TBST and then the Western blot was developed with homemade ECL (enhanced chemiluminescence) reagent.

Samples for semi-quantitative Western blot analysis were prepared similarly to normal Western blot samples with the following modifications. Cells were diluted 1:2,000 and grown at 30 ˚C or 37 ˚C. 5 ml of exponentially growing cells (OD_{600} = 0.2–0.25) were harvested by centrifugation (3,200 g, 10 min, 4 ˚C), the supernatant was discarded and the cell pellet was frozen on dry ice for 15 min. Next, the pellet was resuspended in *x* µl (where *x* is $x = 100 \cdot OD_{600}$) of lysis buffer and incubated for 10 min on ice followed by a 10 min incubation at 37 ˚C on an inverter. The lysis buffer was prepared fresh before use and is composed of B-PER reagent (Thermo Scientific, prod# 78248) with 50 µg/ml lysozyme (Sigma-Aldrich, cat# L6876), Benzonase nuclease (Sigma-Aldrich, cat# E1014, used as a 2,000x stock) and 1x EDTA-free protease inhibitor cocktail (Roche, cat# 11 873 580 001, 1 tablet dissolved in 2 ml H2O was used as a 25x stock). The total

protein concentration of the cell lysate was determined with the Bio-Rad DC Protein Assay (Bio-Rad, cat# 500-0116) according to the manufacturer's instruction manual. The samples were diluted with H_2O and 4x SDS loading buffer to 1x final, boiled for 5 min at 95 ˚C and 10 or 25 µg of total cell lysate was loaded per lane. Transfer, incubation and washing steps were carried out as previously described for standard Western blotting. The primary antibodies used were polyclonal rabbit anti-ClpX and anti-ClpP antibodies (both from the lab collection of Prof. T. Baker, MIT) and were diluted 1:3,000 and 1:5,000 respectively in antibody dilution buffer. After visualization of the ClpX and ClpP bands, the membranes were washed twice for 20 min in 1x TBST, blocked again and re-probed with a mouse monoclonal anti-Sigma70 antibody (Neoclone, cat# WP004) to measure the sigma70 (also known as RpoD) levels as a loading control. The primary anti-Sigma70 antibody was diluted 1:1,000 in antibody dilution buffer and the anti-mouse HRP-linked secondary antibody (GE Healthcare, NA931V) was diluted 1:5,000 in blocking buffer. The Western blot was developed with homemade ECL reagent.

Epi-fluorescence and time-lapse microscopy

All epi-fluorescence microscopy experiments were performed on an inverted microscope (Nikon Ti-E) equipped with a Perfect Focus System (PFS, Nikon), an Orca R2 (Hamamatsu) camera, a Scion CFW-1612M (Scion corporation) camera, a 100x Plan Apo objective (NA 1.4, Nikon), or a 100x total internal reflection fluorescence (TIRF) objective (NA 1.49, Nikon) and an automated x−y stage (BioPrecision2 Inverted Stage, Ludl). The microscope was controlled by Micro-manager (http://www.micro-manager.org/) and custom-written Matlab scripts. For phase-contrast imaging, cells were illuminated with a white LED light source (pE-100, CoolLED), and high-resolution phase-contrast images were captured with the Scion camera (44 nm effective pixel size). Fluorescence imaging was performed with an LED system (Spectra 7 light engine,

Lumencor) and appropriate filter cubes: for DAPI, LF405-A (Semrock); cyan fluorescence, CFP-2432A (Semrock); green fluorescence, GFP-3035B (Semrock); yellow fluorescence, YFP-2427A (Semrock); TMR, TRITC-A (Semrock); and red fluorescence, mCherry-A (Semrock). The fluorescence images were captured nonbinned with the Orca R2 camera (64.5 nm effective pixel size) and saved as 16-bit TIFF images. Overnight liquid cultures, started from a single colony, were grown at 30 °C or 37 °C in a shaking incubator. Cells were diluted, 14–16 h after inoculation, 1:1,000 into imaging medium (M9 medium with 0.2% (w/v) glucose and 10% (v/v) LB medium). Cells were usually grown until early exponential phase (OD600 = 0.1–0.2), diluted with imaging medium, and ~5 µl cell suspension was spotted onto an agar pad. Cells were allowed to sit on the pad for 5–20 min before imaging. The agar pads were made of 2% (w/v) low-gelling agarose (Sigma-Aldrich or QA-Agarose, MP Biomedicals) dissolved in imaging medium. The microscope was equipped with a large incubator, and time-lapse imaging was performed at 30 °C or 37 °C. Coverslips were sequentially sonicated for 30 min in 1 M KOH, acetone, 100% ethanol and double-deionized (dd)H_2O, and then extensively rinsed with ddH_2O before use.

For the fluorescent protein survey, cells were grown at 37 °C to early exponential phase, spotted on an agar pad and then imaged at room temperature (22–26 °C). The strain with the ClpP-mGFPmut3 fusion was grown at 30 °C before imaging. Typical exposure times were 500–2,000 ms for the non-foci-forming fusions and 20–2,000 ms for the foci-forming fluorescent proteins.

Measurement of single-cell degradation rates in daughter cells after cell division

E. coli cultures were grown overnight in LB medium, supplemented with 100 µg/ml ampicillin, at 37 °C in a shaking incubator. Cells were diluted, 14–16 h after inoculation, 1:1,000 into imaging medium (without antibiotics), grown for ~100 min (to 'wake up' from stationary phase), diluted 1:10 and induced for 2 h with 2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) to produce the mCherry-ssrA(LAA) tag degradation reporter. After the IPTG induction, cells were pelleted (4,000 g, 2 min, 37 °C), washed three times with prewarmed imaging medium, diluted 1:10 into prewarmed imaging medium, grown for 15 min (250 rpm, 37 °C), diluted 1:20–1:50 into prewarmed imaging medium, and 2–5 µl cell suspension was spotted on a pre- warmed agar pad. Care was taken to keep the cells at 37 °C during handling and to avoid any temperature changes. The time-lapse experiments were usually started ~40 min after washing away the IPTG. The microscope was enclosed by an in-house built incubator, and imaging was performed at 37 °C (or 30 °C with previous cell growth also at 30 °C and cell handling at room temperature). Five to fifteen stage positions, each having usually one individual *E. coli* cell, were manually selected (only based on the phase- contrast image) and imaged every 5 min in a time-lapse fashion to monitor growth into a micro-colony. A custom-written Matlab script was used to track the micro-colonies during imaging and to correct for x-y drift caused by movement of the agar pad. In short, a 2 x 2 binned phase-contrast image was acquired at every time point, segmented to find the microcolony, and the center of the microcolony was automatically moved to the center of the field of view of the camera before the acquisition of a phase contrast z-stack (5–6 planes with 0.2-µm spacing) and of the fluorescence images. Exposure times for green or yellow fluorescence images were 50–1,000 ms and for red fluorescence images, 200 ms. Red

fluorescence images were taken every time point with green and yellow fluorescence images (to monitor the protease) being less frequent (for example, every fifth time point). Dark field images were acquired for all exposure times and subtracted from the respective fluorescence images.

Quantitative image and data analysis was done using the Schnitzcell program[183,184] (courtesy of M. Elowitz, California Institute of Technology) and Matlab. The *E. coli* cells were segmented based on a 0.2–0.3-μm out-of-focus high-resolution phase-contrast image. Cells were tracked for 30 frames at 30 °C or 20 frames at 37 °C (5 min between frames in both cases). Tracking of cells that grew out of the in-focus monolayer of the micro-colony was aborted, and those cells were excluded from the analysis. Cell segmentation and cell tracking were manually checked, and errors were corrected. Single-cell degradation rates (that is, decrease of total red fluorescence intensity per cell over time) were measured in daughter cells after cell division. The reported degradation rates correspond to the average rates during the cell cycle of the daughter cells (that is, measured from birth to the next division). If one cell divided earlier, the last frame of the earlier dividing daughter cell defined the end point for calculating the average degradation rate of both siblings. Daughter cells with very low red fluorescence (corresponding to cells at the end of the pulse-induction experiment when most of the mCherry-ssrA proteins were degraded) and daughter cells with degradation rates below 10,000 or that had negative values (owing to delayed mCherry maturation at the beginning of the experiment) were excluded from the analysis.

Immunofluorescence microscopy

E. coli cells were grown to early exponential phase (OD_{600} = 0.1–0.15) in LB medium at 30 °C with shaking at 250 rpm. One milliliter of cells was pelleted (4,000 g, 2 min, room temperature), resuspended in 500 μl freshly prepared fixation solution (30 mM sodium

phosphate buffer pH 7.4, 2.5% formaldehyde; Ted Pella) and incubated for 10 min at room temperature on an inverter followed by 30 min on ice. The cells were pelleted (4,000 g, 2 min, 4 °C) and washed 3 times with 500 µl ice-cold 1x PBS. Cells were pelleted again and incubated for 5 min at room temperature in 25 µl GTE buffer (20 mM Tris-HCl pH 7.5, 50 mM glucose and 10 mM EDTA) supplemented with 2.5 µg/ml lysozyme (Sigma-Aldrich). After the incubation, the cells were diluted 1:10 with GTE buffer, and 10 µl of the cell suspension was spread on a KOH-cleaned coverslip, which was coated with 0.01% poly(lysine) solution (Sigma-Aldrich). The cells were allowed to air-dry (> 30 min). Next, 200 µl blocking buffer (1x PBS, 2% (w/v) BSA, 0.05% (v/v) Tween-20) was added to the cover- slip and incubated for 15 min at room temperature. Antibody to ClpX (see below) was diluted 1:100 in blocking buffer, and each coverslip was incubated with 200 µl antibody solution for 1 h in a homemade humidity chamber. The coverslips were incubated 3 times for 5 min with 200 µl blocking buffer. Secondary antibody (Alexa Fluor 350 donkey anti-rabbit IgG (Invitrogen, A10039)) was diluted 1:500 in blocking buffer and ~300 µl was added to each coverslip. The coverslips were incubated in the dark for 1 h. Next, cells were washed three times with 200 µl blocking buffer with 5-min incubations between washes. Coverslips were mounted on agar pads made of 2% low-gelling temperature agarose (Sigma-Aldrich) in 1x PBS. The immunofluorescence samples were imaged on an inverted microscope (Nikon Ti-E) equipped with a 100x TIRF objective (NA 1.49, Nikon), an Orca-R2 camera (Hamamatsu) and a shuttered mercury light source (Intensilight, Nikon). The Alexa Fluor 350 fluorophores were imaged with a DAPI (LF405-A, Semrock) filter cube.

Antibodies to ClpX

Antibodies to ClpX were purified using Affigel-10 (Bio-Rad) resin. The *E. coli* ClpX protein was purified according to previous protocols[95] and conjugated to resin; ClpX

polyclonal rabbit antibodies (Covance) were purified according to the manufacturer's protocol. Affinity-purified antibodies were aliquoted in 1x PBS containing 50% (v/v) glycerol and stored at –80 °C.

SNAP tag labeling in fixed *E. coli* cells

The SNAP tag has not previously been used to detect endogenous intracellular proteins in bacteria to our knowledge, and hence the existing protocols needed to be carefully optimized. The ClpP-SNAP tag-expressing strain and the wild-type strain, which served as a negative control for the specificity of the labeling reaction, were processed in parallel and subjected to identical treatments. First, *E. coli* cells were grown to mid-exponential phase in LB medium at 37 °C with shaking (250 rpm). Cells were fixed with 2.5% (v/v) formaldehyde by adding 1.56 ml of a 16% formaldehyde solution (Ted Pella) directly to 8.44 ml cell culture and incubated at room temperature for 30 min in a shaking incubator (250 rpm). After fixation, the cell suspension was pelleted by centrifugation (4,000 g, 10 min, 4 °C), washed with 2 M glycine to quench the fixation reaction and incubated in 1 ml 2 M glycine for 1 h in a thermomixer (Eppendorf) at room temperature with 1,400 rpm shaking. The 1 h incubation with 2 M glycine reduced the nonspecific binding of the TMR SNAP-Cell dye[185]. The SNAP-Cell TMR dye (NEB) was diluted in DMSO (Sigma-Aldrich) to a final concentration of 30 µM and was used as a 100x stock. After the 1 h glycine incubation, cells were pelleted (4,000 g, 2 min, room temperature), washed with 1 ml washing buffer (1x PBS, 200 mM glycine, 1 mM DTT, 2% (w/v) BSA and 0.05% (v/v) Tween-20) and incubated in the thermomixer for 10 min (1,400 rpm, room temperature). Cells were then pelleted again and concentrated in 100 µl washing buffer, followed by the addition of 1 µl of 30 µM SNAP-Cell TMR dye and incubated for 30 min in the thermomixer (1,400 rpm, room temperature). The labeling reaction and all the following steps were done under low-light conditions. After incubation with the TMR

dye, cells were subjected to excessive washing steps to remove nonspecifically bound dye molecules. The cells were pelleted (4,000 g, 2 min, room temperature) and were washed ~10 times with 1 ml wash buffer and incubated for ~1 h between washes in a thermomixer (1,400 rpm, room temperature). The progression of the washing was monitored by observing the sample and the negative control side by side. We found out later that shorter washing steps (e.g. five 5–10-min washes in the thermomixer, 1,400 rpm, room temperature) and using a new Eppendorf tube after every wash dramatically increased the washing efficiency and hence making the 1 h washing steps obsolete.

Before microscopy, cells were washed twice with 1x PBS, diluted with 1x PBS (if necessary) and imaged on an agarose pad (2% (w/v) low-gelling agarose dissolved in 1x PBS) or squashed between a coverslip, coated with 0.01% poly(L-lysine) (Sigma-Aldrich), and a microscope slide.

Imaging was performed with an inverted microscope (Nikon Ti-E) equipped with an Orca R2 (Hamamatsu) camera, a Spectra 7 light engine (Lumencor) and a 100x Plan Apo objective (NA 1.4, Nikon). The TMR-stained bacteria were imaged with the green LED (excitation filter 549/15 nm, 40 mW power measured out of the objective) and a Cy3 filter cube (TRITC-A, Semrock) with typical exposure times of 2–5 s.

Highly inclined and laminated optical sheet (HILO) microscopy

Cells were grown at 30 °C or 37 °C to exponential phase and the microscopy was performed at room temperature. Bacteria were imaged via an in-house built objective-type TIRF-based configuration on a modified inverted microscope (Nikon Ti-E) equipped with a 100x Plan Apo objective (NA 1.4, Nikon) and a 2.5x C-mount adaptor (Nikon). Illumination was achieved by a 488-nm (Sapphire, Coherent) laser line, which was focused on the back focal plane of the objective. As the TIRF modality does not allow the whole bacterial cell volume to be illuminated, the laser light was diverted from TIRF

and sent to the sample at an inclined fashion[131], that is, with an angle that is slightly steeper than the critical angle. The incidence angle was adjusted manually by varying the position of the focusing lens sitting on an xyz stage (Newport) and was empirically optimized for the best signal-to-noise ratio. Images were acquired using an electron-multiplying charge-coupled device (EMCCD) camera (ixon3 897, Andor) with EM gain set to 300 and a 5x pre-amplifier gain. The camera was controlled with the software Solis (Andor). The effective pixel size of the acquired images corresponded to 64 nm (with 250x magnification). To acquire fast dynamics, only a subregion of the EMCCD chip was used. ClpA-mGFPmut3, ClpX-mGFPmut3, ClpP-mGFPmut3 and mGFPmut3 alone live-cell movies were acquired under constant illumination with 10 mW laser intensity and 30 Hz image acquisition rate. For green fluorescence imaging, a dichroic filter (Di01-R488, Semrock) and a bandpass emission filter (FF01-525/45, Semrock) were mounted on a custom-made brass filter cube.

The image sequences were exported from Solis as 16-bit TIFF files and cropped to 128 x 128 pixel in ImageJ (http://rsbweb.nih.gov/ij/). The images were then loaded into Matlab for batch processing and intensity scaled. For quantitative intensity scaling, all images in a movie sequence had the same grayscale scaling, where the pixel with the highest numerical value in the image sequence corresponded to the white pixel (1) and the pixel with the lowest numerical value to the black pixel (0). For 'per-frame autoscaling', every image in the sequence was subjected to an individual 'min-max' scaling, where the brightest pixel in the individual image corresponded to the white pixel (1) and the pixel with the lowest numerical value to the black pixel (0). The resulting images were loaded into QuickTime 7 Pro (version 7.6.6, Apple) and compressed with the H.264 codec.

Live *E. coli* cells expressing the ClpP-Dronpa fusion (strain DHL726) were imaged at room temperature under constant illumination with a 488 nm laser (50 mW) in HILO

mode and a time-series of 200 frames (30 ms exposure time, 250x magnification, no binning, 64 nm effective pixel size) was recorded. The Dronpa molecules were not photo-converted with a 405 nm laser during the experiment. The time-series was loaded into imageJ, cropped and converted into a maximum projection.

The images for the live-cell HILO time-series (Figure 2.17) were acquired similarly except with 150x magnification and a Hamamatsu EMCCD camera (C9100-02) with an effective pixel size of 53.3 nm and 50 EM gain. Images in sequence were acquired every second with 20 ms exposure time and shuttered 15 mW 488-nm laser illumination in HILO mode.

Short fixation of cells containing the ClpP-mGFPmut3 or ClpX-mGFPmut3 fusion

Exponential-phase cells, harboring the ClpP-mGFPmut3 or ClpX-mGFPmut3 fusion, were grown in imaging medium, spun down (4,000 g, 2 min, 4 °C) and resuspended in 1 ml fixation solution (2.5% formaldehyde, 30 mM sodium phosphate buffer pH 7.4) and incubated for 5 min on ice. Cells were pelleted and washed with 1 M glycine, followed by three washes with 1x PBS. Finally, the cells were squashed between a KOH-cleaned coverslip and a microscope slide and imaged immediately with the same setup that was used for live-cell HILO microscopy. Image sequences were acquired with constant illumination using 30 ms integration time and 30 mW (ClpX-mGFPmut3, Figure 2.16, left) or 10 mW (ClpP-mGFPmut3, Figure 2.16, right) 488-nm laser intensity. The EM gain of the EMCCD camera was set to 50, and 5x pre-amplifier gain was used. The TIFF image sequences were loaded into ImageJ to generate sum projections of the first 100 (ClpX-mGFPmut3) or 300 (ClpP- mGFPmut3) frames.

2.5. Description of Movies

Video 2.1

Micro-colony growth of *E. coli* cells with ClpP-Venus foci. ClpP-Venus formed bright fluorescent foci in live *E. coli* cells. The ClpP-Venus foci were not present in all cells, the foci localized preferentially to cell poles and mid-cell region and showed binary segregation at cell division. Cells without a focus had a cytoplasmic yellow fluorescence signal (not visible in this movie) and usually formed a yellow fluorescence focus in the next cell cycles. A few ClpP-Venus foci appeared blurred because they were in different focal planes (z-stacks were required to detect all foci in a microcolony; data not shown). Some cells had small foci, which are very faint and barely visible in the movie. The yellow fluorescence images of this movie were acquired with 2 x 2 binning (effective pixel size was 129 nm). The cell boundary (red) was determined by segmenting the respective phase-contrast images. Scale bar (white), 1 µm. The movie is part of a dual-color experiment; see **Video 2.2** for the corresponding red fluorescence movie.

Video 2.2

Degradation of mCherry-ssrA in the ClpP-Venus strain showed that the ClpP-Venus foci generate cell-to-cell variability after division. Synthesis of mCherry-ssrA was pulse-induced before the cells were monitored. The red fluorescence images were subjected to a 'per-frame auto-scaling' to better display the variability between daughter cells after division. The foci in the red fluorescence image were not due to spectral bleedthrough from ClpP-Venus foci but probably represent immortal mCherry molecules bound to the ClpP-fluorescent protein foci. Red fluorescence images of this movie were acquired with

2 x 2 binning (effective pixel size was 129 nm). The cell boundary (red) was determined by segmenting the respective phase-contrast images. Scale bar (white), 1 µm.

Video 2.3

Degradation of mCherry-ssrA in the wild-type strain (protease was not tagged) displayed very low cell-to-cell variability after cell division. The mCherry-ssrA reporter was pulse-induced before imaging. mCherry-ssrA is specifically degraded by ClpXP and ClpAP. The red fluorescence images were per-frame auto-scaled to better illustrate the low variability after cell division. The red fluorescence images of this movie were acquired with 2 x 2 binning (129 nm effective pixel size). The cell boundary (red) was determined by segmenting the respective phase images. Scale bar (white), 1 µm.

Video 2.4–2.6

Live-cell HILO microscopy of cells expressing ClpA-mGFPmut3 (**Video 2.4**), ClpP-mGFPmut3 (**Video 2.5**) and ClpX-mGFPmut3 (**Video 2.6**). Cells have ~50 particles, which are not localized in foci but move rapidly and seem to sample the entire cell. Scale bars (white), 1 µm. The images of the movie sequences were subjected to a quantitative grayscale scaling (left) and to a 'per-frame auto-scaling' (right) to better display the particle movement despite fast photo-bleaching.

Video 2.7

Live-cell HILO microscopy of cells expressing mGFPmut3 alone (that is, not fused to another protein) expressed from the clpX promoter at the endogenous locus. The fluorescence signal displays a uniform cytoplasmic distribution. Scale bars (white), 1 µm. The images of the movie sequences were subjected to a quantitative grayscale scaling (left) and to a 'per-frame auto-scaling' (right).

Video 2.8

Micro-colony growth of *E. coli* cells with Hfq-mGFPmut3, PepP-mGFPmut3, IbpA-mGFPmut3, MviM-mGFPmut3 and FruK-mGFPmut3 fusions. The fusions are constructed at the endogenous gene loci. Cells were grown to exponential phase in imaging medium and micro-colony growth was filmed on an agar pad with exposures every 10 min. Cell growth and imaging was performed at 30 °C. Green fluorescence images with 1-s exposure time were acquired every 10 min. Green fluorescence images of the movies were acquired with no binning (effective pixel size is 64.5 nm) and were subjected to a quantitative grayscale scaling. Scale bars (white), 1 μm.

Counting mRNA molecules simultaneously with two single-molecule methods

3.1. Introduction

3.1.1. Reliable mRNA counting in individual cells

Many mRNAs are present in very low numbers per cell. Whether analyzing how gene expression is adjusted in response to internal changes in the cell, or studying how random fluctuations originate and spread, this requires methods to reliably measure single-cell mRNA levels. In fact, many studies of non-genetic heterogeneity in protein levels[21,186-194] concluded that the observed heterogeneity is largely caused by low abundance mRNAs, but often reached this conclusion indirectly by fitting the properties of protein fluctuations to stochastic models rather than actually observing mRNA fluctuations.

In principle, individual mRNA molecules can be detected and quantified in individual cells with single-molecule fluorescence *in situ* hybridization (FISH)[33,195,196], without modifying the mRNA target or any need for time-consuming cloning. This method has single-molecule sensitivity in the sense that it detects single mRNA molecules. It is therefore often assumed that FISH can be used to count the number of molecules per cell, but counting accuracy has to our knowledge never been demonstrated. Another approach uses RNA-binding proteins fused to fluorescent proteins, again to visually identify mRNA as fluorescent foci in cells. This approach has been used with single-

molecule accuracy[32], but only after inserting 96 protein-binding sites onto the mRNA, which changes mRNA levels and many of its properties.

The challenge of developing mRNA counting methods is to not only develop a method that works accurately, but a method that *provably* works accurately. Unfortunately there is no gold standard where the exact number of mRNA molecules per cell is known. Without such a standard, elaborate internal controls are required to demonstrate that both false-positive counts and false-negative counts are negligible.

This chapter focuses on our attempts to develop a reliable method for mRNA counting in live and fixed *Escherichia coli* (*E. coli*) cells. Sections 3.1.2 and 3.1.3 discuss previous approaches in the literature, focusing on the challenges they face and the experimental controls presented, while Section 3.1.4 outlines our strategy. Section 3.2.1 then describes the design and internal evaluations of our FISH-based methods, Section 3.2.2 does the same for RNA-binding proteins, and 3.2.3 shows some preliminary results for cross-validation by detecting the same individual transcripts with both methods simultaneously. Currently our results suggest that the existing mRNA counting assays indeed are not very quantitative, but also identify several problems that could be straightforward to resolve, possibly resulting in an independently validated quantitative assay. This is work in progress and if high accuracy can be established, we will use it to count mRNA molecules of stress response genes (relevant to work in Chapter 4).

3.1.2. Existing FISH assays for mRNA counting in fixed cells

Robert Singer's laboratory demonstrated in 1998 that single β-actin mRNA molecules could be detected in fixed mammalian cells by using several FISH probes that were each labeled with five dyes per probe and were targeted specifically to the mRNA of

interest[195]. Binding of multiple probes resulted in the formation of diffraction-limited point sources, which could be detected by fluorescence microscopy. Counting the number of diffraction-limited spots above a defined threshold would result in the total number of mRNA transcripts per cell. Messenger RNAs that are closer together than 200 nm (the diffraction limit) would not be spatially resolved and could appear as one object. This would result in under-counting of mRNAs but unless the molecules cluster, it should not be a substantial problem for low-copy transcripts in large cells.

The Singer lab partially resolved this issue when they extended single-molecule mRNA FISH to smaller cells (i.e. budding yeast)[33]: Zenklusen and colleagues[33] showed that the integrated fluorescent intensity of a diffraction-limited spot contains information about the number of mRNA molecules that are contained within the spot. Their statistical analysis included a histogram of ~1000 spots but the histogram is heavily smeared out and did not produce equidistant peaks corresponding to the number for mRNA molecules per spot. Gaussian fitting was necessary to estimate the number of transcripts per spot. The poorly fitted Gaussian peaks were rather broad and overlapped, and yet this is the most quantitative data we have seen reported in the FISH literature. Intensity-based corrections are attractive because they could be used to obtain information beyond the diffraction limit, which would be even more important for smaller cells like *E. coli* that have a ~20x smaller volume than haploid yeast cells.

A second alternative FISH method uses a large number (usually > 48) of singly labeled probes[196,197]. The advantage of this method is that the probes are easier to synthesize, label and purify. Furthermore, short probes may penetrate better into cells (particularly into tissue) since they are smaller in size. A potential problem of this approach, however, is that the short probes have less specificity and so will likely bind to more off-target sites. Also, the probes cannot practically be evaluated individually since there are too many of them, and non-specifically bound probes will increase the fluorescent

background. However, the reduced specificity is compensated by a larger number of probes. We expect that this will increase the signal-to-background ratio, but that partial specific binding combined with non-specific binding decrease the signal-to-noise ratio in the intensity per spot. This approach may therefore be less appropriate for inferring the number of mRNAs within a diffraction-limited spot by relying on the fluorescent intensities. Two recent studies[198,199] did use it in *E. coli* and based the counts almost exclusively on total intensities. However, these studies focused on higher abundant mRNAs and included no accuracy controls or statistics supporting the fact that the method worked with anything near single-molecule accuracy. For higher organisms where overlap of diffraction-limited spots is not an issue, this is less of a problem, and signal-to-background may be more important. Currently, the short probes are also commercially available at a relatively low cost, which is perhaps one reason why this method has become so common and applied to a large number of different specimen, including whole organisms[200], mammalian cells[196], yeast[196] and *E. coli*[198,199].

A third type of FISH assay was developed in a recent high-throughput *E. coli* study that reported global mRNA and protein quantification[21]. The study surprisingly found no correlation between the mRNA and protein levels in single cells for over 100 highly expressed genes that were analyzed. The mRNA FISH method relied on a single but highly optimized probe. Non-specific binding events to off-targets then cannot be distinguished from real events and would result in over-counting. The washing steps of the FISH protocol are also often optimized to ensure that probes do not bind in a strain where the mRNA target has been deleted. For a single probe, this optimization protocol may greatly increase the number of false negatives (mRNAs that are not bound by a probe) and lead to under-counting than decrease the number of false positives (non-specifically bound probes).

The one-probe FISH method is likely the least quantitative of the three available approaches, and is also virtually impossible to validate with internal controls. The other two methods have potential advantages and disadvantages depending on the type of cells studied. A recent study also multiplexed detection of mRNA molecules in *Saccharomyces cerevisiae* with spectrally barcoded FISH probes and stochastic optical reconstruction microscopy (STORM)[201]. This approach is very promising but also did not prove single-molecule counting accuracy. The method is further sensitive to barcode assignment errors and suffers from most of the traditional single-molecule FISH problems like probe hybridization efficiency, quantitative and specific binding, and probe penetration issues as discussed here and in a recent review about plasmid quantification in single bacterial cells[202].

3.1.3. Existing reporters for mRNA detection and counting in live cells

Individual mRNA molecules can also be detected and quantified in living cells. Virtually all current approaches for live-cell mRNA detection require the mRNA to be modified by adding a tag to the beginning or the end of the message. The addition of a tag can change mRNA movement[32], mRNA levels by preventing degradation[32] and may cause mRNA mislocalization. Appropriate controls are therefore necessary to show that the tag is non-intrusive.

The most widely used live-cell mRNA imaging reporter is the MS2 system[203], where the mRNA of interest is tagged with an array containing multiple MS2-binding sites. Co-expressing the cognate RNA-binding protein from bacteriophage MS2 fused to a fluorescent protein then allows detection of the mRNA *in vivo*. The native function of the MS2 protein is to form the virion shell of the bacteriophage MS2, which is composed of 180 MS2 molecules. It is therefore unsurprising that the MS2 protein is prone to aggregation[204] and focus formation[203], likely due to strong avidity effects caused by the

MS2 oligomers[205]. Less aggregation-sensitive versions of MS2 exist[204] but are rarely used. Despite this disadvantage the MS2 system was used to accurately count single mRNA molecules in live *E. coli* cells[32]. The mRNA molecules were tagged with 96 MS2-binding sites and co-expression of a GFP-tagged MS2 fusion drove individual mRNA transcripts usually into a single diffraction-limited focus per cell. A histogram of focus intensities revealed multiple equidistant peaks, which were independently shown to correspond to integer numbers of mRNA molecules. Combined with several other accuracy controls, this suggested that the approach had close to single molecule resolution. However, having 96 repeats with many bound MS2-GFP proteins[144] completely prevented degradation (which was exploited in the study to analyze transcription) and is expected to drastically affect localization and movement, while shorter binding site repeats did not have single-molecule resolution. A two-binding site MS2 tag was also used in *E. coli* to study mRNA diffusion with fluorescence correlation spectroscopy (FCS)[206,207]. Quantification of mRNA levels should also be possible with FCS, since FCS is inherently quantitative, but probably not with single-molecule accuracy.

The RNA-binding protein PP7 of the *Pseudomonas aeroginosa* phage is a distant relative of MS2 and also binds to a short RNA hairpin (**Figure 3.1**).

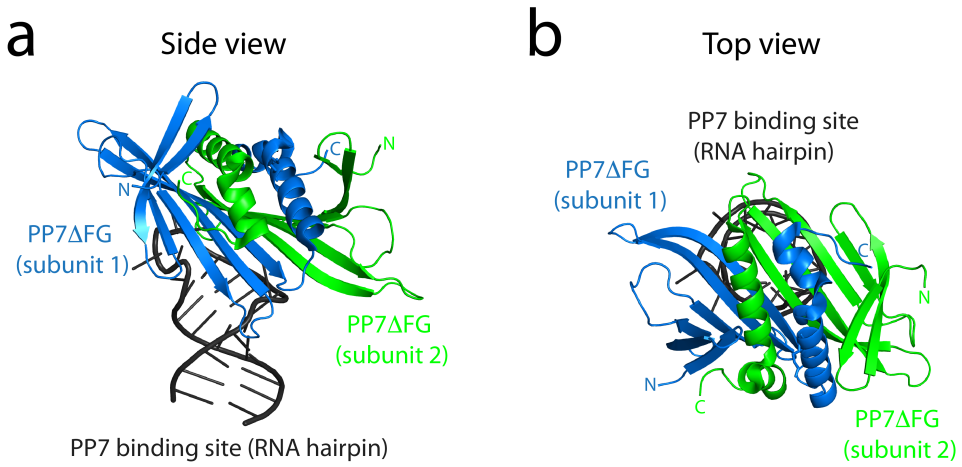


Figure 3.1. Two PP7ΔFG molecules (blue and green) dimerize and bind to an RNA hairpin (black). Side view (**a**) and top view (**b**) are shown. The structure of the complex[208] was solved by X-ray crystallography (PDB entry 2QUX).

Similar to MS2, the PP7 system has been used for live-cell mRNA imaging[209]. A recent comparative study[145] of MS2ΔFG and PP7ΔFG (the aggregation-deficient versions) found that PP7ΔFG is better for mRNA imaging since it binds its respective hairpin at lower concentrations (smaller dissociation constant), which allows imaging at lower cytoplasmic background levels, and that PP7ΔFG binding reaches full occupancy of the array (whereas MS2ΔFG reaches only ~50%). The study also suggested using tandem dimers of the coat proteins (i.e. tdPP7ΔFG) since this eliminates the dimer formation step (the dimer binds to the hairpin) and hence reduces the fluorescence background even further[145].

Finally, an RNA aptamer, dubbed Spinach[210], that becomes fluorescent upon binding of a non-fluorescent small molecule was recently developed and used for mRNA imaging in mammalian cells and also bacteria, though all imaged RNAs were extremely highly expressed. Our experiments suggested that this approach is extremely inaccurate and our results are therefore only included in the appendix (Section 3.5).

3.1.4. Our approach

We hope to develop an mRNA counting method that not only works accurately in *E. coli*, but where the accuracy can be independently validated. To this end, we developed both FISH and PP7 approaches and then evaluate them against each other by testing for co-localization on the same individual transcript molecules.

Specifically, FISH can produce false positives due to non-specific binding of the probes or poor washing, although evaluating the assay against a knockout strain that does not produce the target can easily test this. False negatives or heterogeneity in binding are harder to evaluate, and could be caused by incomplete labeling of the probes, by partial probe penetration into the cells, by incomplete binding to the target, or by several other problems due to diffraction-limited spots or non-quantitative microscopy. We first evaluated the full method as described in the literature, to determine overall accuracy. We then analyzed the various steps using a combination of *in vitro* and *in vivo* methods, from counting photo-bleaching steps of the probes, to histograms of spot intensities.

We similarly evaluated many of the steps of the PP7 assay, and finally brought the two together to check for co-localization of target molecules with the two types of tags. The two methods thus evaluate each other's rates of producing false negatives. Such controls cannot be performed by merely looking for co-localization between two differently colored FISH probes. For example, if some mRNAs are temporarily inaccessible due to fixation, the probes will still co-localize and fail to report the false negatives.

Our dual-labeling experiments were so far performed *in vitro* only but will be extended to fixed *E. coli* cells to detect the transcripts *in situ*. In collaboration with Maier Avendano, Ralf Jungmann and Peng Yin, we are also attempting to extend mRNA FISH for counting high-copy mRNAs and multiple mRNA species in parallel by using DNA-

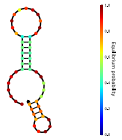
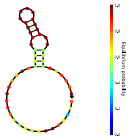
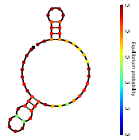
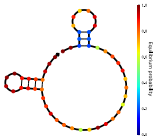
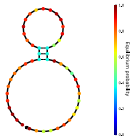
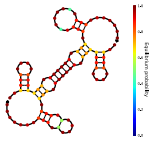
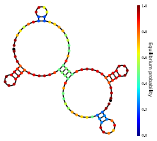
PAINT[211], which will permit sub-diffraction quantification of endogenous biomolecules in single cells (data not presented here).

3.2. Results

3.2.1. FISH for detecting *clpX* mRNA transcripts

I chose the *clpX* mRNA as a model mRNA to establish the FISH method partly since this mRNA appears to be present in relatively low numbers per cell (5–10 copies) and partly because we are interested in the ClpXP proteolysis system. I designed 50-nucleotide (nt) long FISH probes against the *clpX* mRNA coding region following guidelines from the Singer lab[212]. A pool of 135 candidate FISH probes was generated and the probes were further computationally analyzed for secondary structure formation, dimerization and off-target binding sites (Materials and Methods). The FISH probes used are listed in **Table 3.1**.

Table 3.1. Chemical properties and secondary structures of the *clpX* FISH probes

Probe	***clpX*-77**	***clpX*-433**	***cpX*-674**	***clpX*-853**	***clpX*-999**
Length GC content Melting temp. $\Delta G_{worst\ hairpin}$ $\Delta G_{worst\ self\text{-}dimer}$ (Oligo Analyzer 3.1)	50 bp 50% 76.1 ˚C –2.92 kcal/mol –9.75 kcal/mol	50 bp 50% 76.0 ˚C –2.15 kcal/mol –9.75 kcal/mol	50 bp 50% 76.9 ˚C –2.92 kcal/mol –8.51 kcal/mol	50 bp 50% 76.1 ˚C –2.17 kcal/mol –9.89 kcal/mol	50 bp 50% 76.0 ˚C –0.91 kcal/mol –9.89 kcal/mol
Hairpin formation (NUPACK)	 ΔG = –4.97 kcal/mol	 ΔG = –4.12 kcal/mol	 ΔG = –5.06 kcal/mol	 ΔG = –3.66 kcal/mol	 ΔG = –1.70 kcal/mol
Self-dimer formation (NUPACK)	–	–	 ΔG = –18.85 kcal/mol, 1.37% of the molecules form dimers at 1 μM conc.	 ΔG = –15.39 kcal/mol, 0.31% of the molecules form dimers at 1 μM conc.	–
Off-targets (other *E. coli* genes with exact matches)	*yeeS* (14/50) *sapF* (14/50) *malE* (13/50)	*aaeR* (15/50) *cusR* (18/50) *trpE* (14/50)	*baeS* (15/50) *yraR* (15/50) *malE* (14/50)	*wcaK* (16/50) *tauA* (16/50) *hsdS* (1550)	*glnG* (16/50) *mdtC* (16/50) *tamB* (15/50)

The FISH probes were then used to detect *clpX* mRNA transcripts in fixed *E. coli* cells. First we demonstrated a negligible rate of false positives: the *clpX* FISH probes are specific to their target mRNA as demonstrated by their lack of substantial binding to the Δ*clpX* deletion strain. The few observed spots in the knockout strain have intensities corresponding to single or at the most two probes, which probably co-localize by chance or due to a weak affinity (**Figure 3.2a** and **c**). By contrast, bright spots were observed for the wild-type sample and quantification of the spot intensities suggests that those spots correspond to single mRNAs or in some cases multiple mRNAs (**Figure 3.2b** and **d**). The brightest spots that may correspond to 2 or 3 mRNAs could potentially correspond

to mRNA clusters or to the transcriptional start site at the *clpX* gene locus. It was previously suggested[213] that *E. coli* mRNAs co-localize with their cognate gene loci and do not diffuse away after transcription termination. However, preliminary data suggest that the number of mRNA spots observed, and their cellular localization, appeared too high and too dispersed to be consistent with this model.

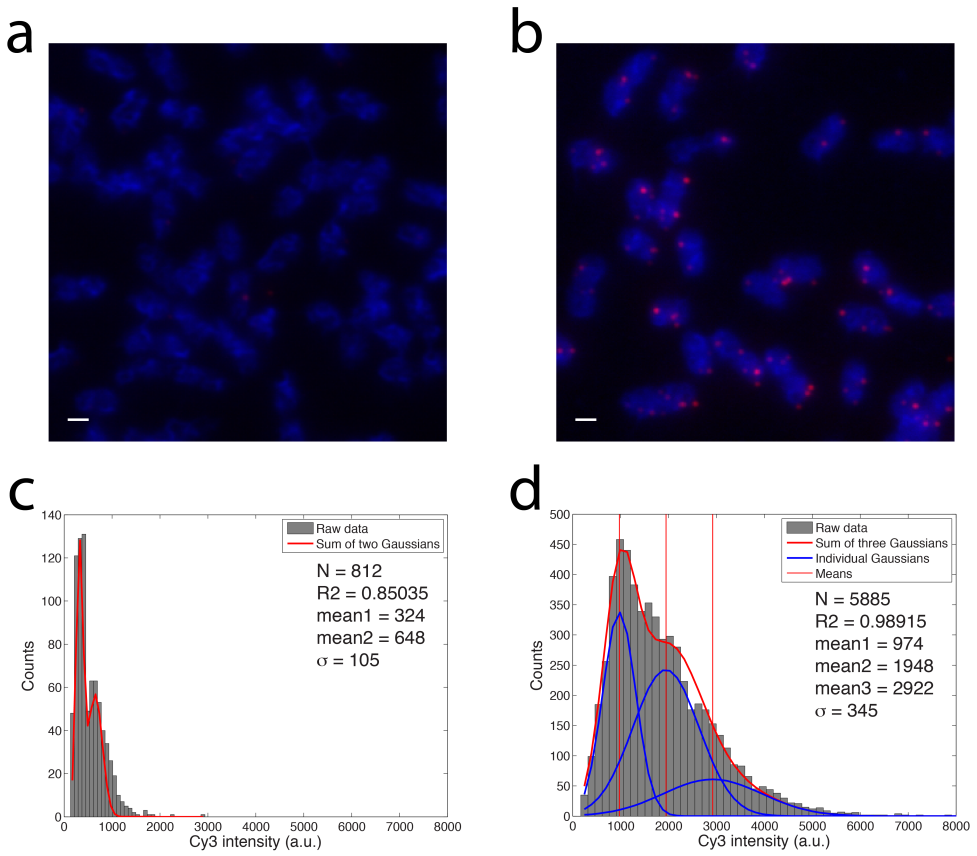


Figure 3.2. Visualization of individual *clpX* mRNA molecules in fixed cells using four *clpX* Cy3 FISH probes (with 5 Cy3 dyes per probe). (**a**) Only a few dim spots are observed in the Δ*clpX* strain. (**b**) Many bright spots were observed in the wild type. DAPI stain (blue) and Cy3 stain (red). Scale bars (white) are 1 μm. (**c**) Quantification of the spots of the Δ*clpX* strain indicates that the spots correspond to one non-specifically bound FISH probe (mean1 = 324) or two probes (mean2 = 648). (**d**) Quantification of the spots in the wild-type strain suggests that the spots correspond to 1, 2 or 3 mRNAs (means = $\boldsymbol{n \cdot 974}$ where $\boldsymbol{n = 1, 2 \text{ or } 3}$), where each mRNA is in average bound by three probes out of a maximum of 4 probes (974/324 = 3.01 probes). The fractional contribution of the areas under the Gaussian curves are Area1 = 34.6% (1 mRNA), Area2 = 52.5% (2 mRNAs) and Area3 = 12.9% (3 mRNAs).

The histogram of spot intensities for the wild-type strain does not show discrete peaks, and the presence of multiple subpopulations (1, 2 and 3 mRNAs) could only be suggested by constrained Gaussian fitting (Materials and Methods). The low separation between the mRNA peaks could be due to a large variability in probe binding, measurement and analysis noise, or incomplete labeling of the FISH probes.

To further investigate the possibility of incomplete probe labeling, I determined the labeling efficiency of the *clpX* Cy3 FISH probes. The probes were non-specifically attached to a poly-lysine-coated coverslip and imaged with an epi-fluorescence microscope. Image analysis of spot intensities (and the number of photo-bleaching steps) revealed that labeling of the probes is incomplete: rather than the five Cy3 dyes that these FISH probes (Materials and Methods) were supposed to have, the probes on average had ~2.1 dyes attached (**Figure 3.3a** and **b**). The reduced number of active dyes could be due to incomplete labeling after probe synthesis (relying on less efficient NHS chemistry) or due to probe decay from multiple freeze-thaw cycles, although this seems less likely.

To rule out inefficient probe labeling, we obtained a new set of *clpX* FISH probes, where each probe was labeled co-synthetically with three Quasar570 dyes and the labeling efficiency is expected to be close to 100%. The biotinylated *clpX-77* Quasar570 probe was surface-immobilized using BSA-biotin and streptavidin and then imaged. The single-molecule imaging test revealed that this probe is also incompletely labeled, although the labeling efficiency appears higher compared to the earlier probe set (**Figure 3.3c** and **d**). The apparent incomplete probe labeling is possibly due to insufficient purification of the probes after synthesis (personal communication with the manufacturer). The manufacturer is currently improving the probe purification by using dual-HPLC after synthesis of the FISH probes. We have also begun analyzing the probe binding

efficiency, but all such tests hinge upon using fully labeled probes: partially labeled probes do not merely reduce the average fluorescence, but also introduces heterogeneity in intensity since some probes have more dyes bound than others. This heterogeneity can easily be mistaken for partial probe binding to the target.

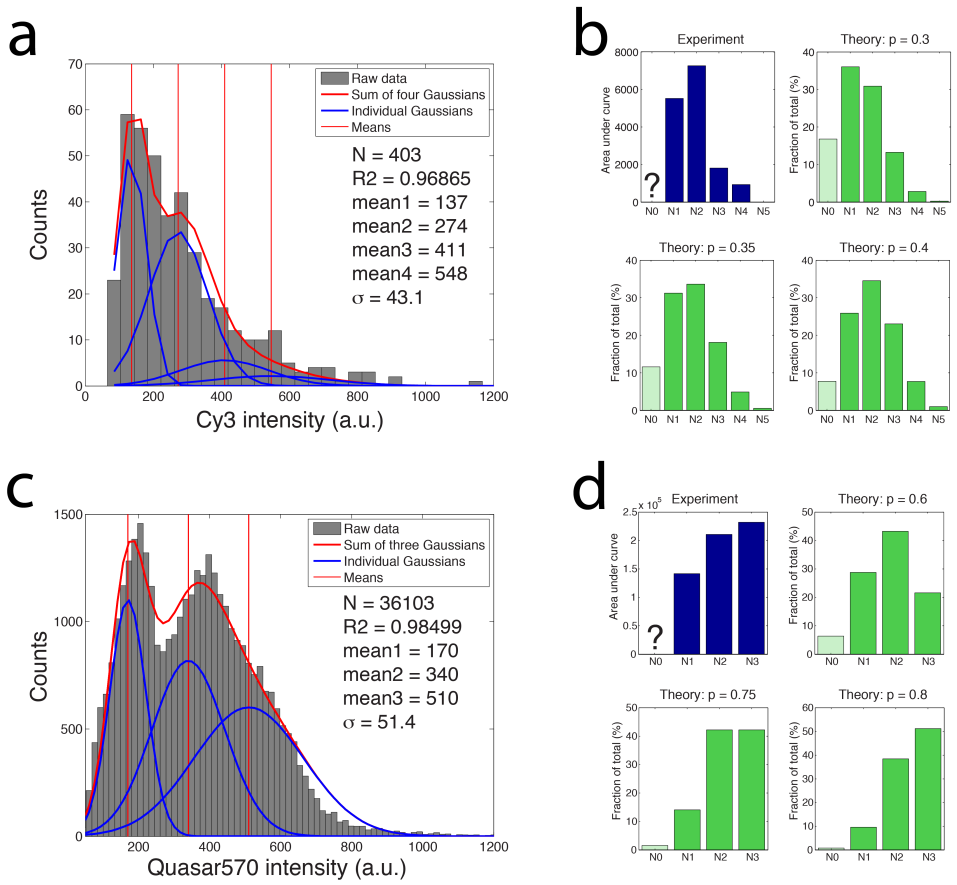


Figure 3.3. All *clpX* FISH probes analyzed were incompletely labeled. (**a**) The *clpX* Cy3 FISH probes are supposed to have 5 dyes but the number of active dyes is much lower. The histogram was fitted with the sum of four Gaussians (Materials and Methods) corresponding to the single, double, triple and quadruple labeled species. Individual Gaussians (blue) have a mean of $x \cdot 137$ and standard deviation of $x \cdot 43.1$ where x is 1, 2, 3 or 4. (**b**) Estimation of the labeling efficiency. The areas under the Gaussian curves (blue bar plot) is compared to the theoretical prediction (green bar plots) of labeling efficiencies of p = 0.3, 0.35 or 0.4 (p is the probability that a site has a fluorescent dye). (**c**) The *clpX*-77 FISH probe is supposed to have three Quasar570 (a Cy3 derivative) dyes. Quantification of 36103 diffraction-limited spots resulted in a histogram with three peaks, which correspond to the single, double and triple labeled probe species. The histogram was fitted with the sum of three Gaussians (Materials and Methods). Individual Gaussians (blue) have a mean of $x \cdot 170$ and standard deviation of $x \cdot 51.4$ where x is 1, 2 or 3. (**d**) The labeling efficiency is probably ~60% (though the fit overestimates the area of the third Gaussian). N0 is the unlabeled probe and hence undetectable (faint green); N1–N5 correspond to the FISH probe with 1 to 5 dyes respectively. Incomplete labeling was also observed when analyzing photobleaching steps (data not shown).

3.2.2. PP7 array for detecting *clpX* mRNA transcripts

An array of 24 PP7 binding sites (referred to herein as the 24xPP7 array) was designed with the software package NUPACK (Materials and Methods). The software allows for designing DNA sequences that adopt a given secondary structure. We used the software to design an array that is composed of 24 well-folded PP7 hairpins (25 nt) that are separated by 20-nt long spacers, which (i) do not interfere with folding of the PP7 hairpins and (ii) have little sequence homology with each other (**Figure 3.4**). Repetitive DNA sequences are very unstable in *E. coli* because of RecA-mediated recombination[214] and also hamper PCR amplification, which complicates the construction of strains with tagged mRNAs, and *in vitro* transcription. Random spacer sequences were previously used for e.g. the 96xMS2 array[32] but our approach has the additional benefit that the intervening sequences are very unlikely to interfere with the folding of the hairpins. Computational analysis with NUPACK suggests that totally random spacers sequences are indeed very likely to interfere with folding of the hairpins.

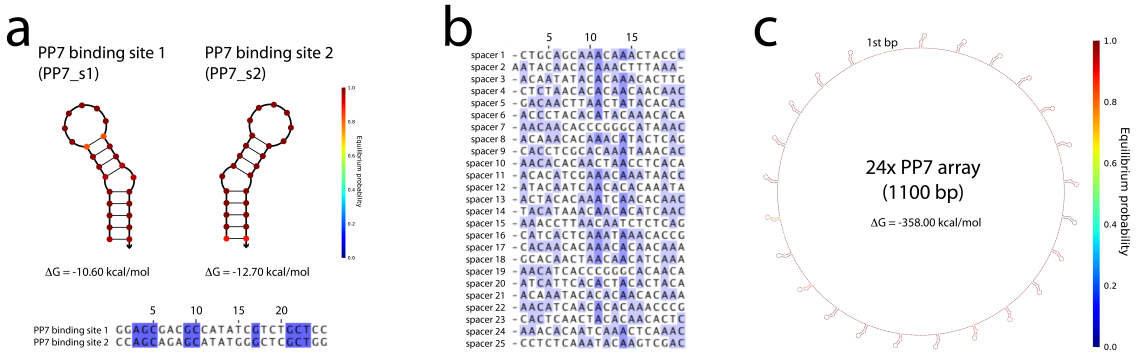


Figure 3.4. Design of the 24xPP7 array. (**a**) The 24xPP7 array is composed of two alternating PP7 binding sites (site 1 and site 2) which share low sequence similarity (9/25). Secondary structures of the two PP7 binding sites were calculated with NUPACK. (**b**) Individual PP7 binding sites are separated by 20-nt long inert spacers (see Materials and Methods) and the spacers have little sequence similarity. (**c**) Simulation with NUPACK predicts that the full-length (1100 bp) 24xPP7 array forms all 24 PP7 hairpins.

The DNA sequence of the 24xPP7 array is given in **Table 3.2**.

Table 3.2

Name	Sequence
24xPP7 array	ctgcagcaaacaaactacccggagcagacgatatggcgtcgctccaatacaacacaaactttaaaccag cagagcatatgggctcgctggacaatatacacaaacacttgggagcagacgatatggcgtcgctccctcta acacacaacaacaacccagcagagcatatgggctcgctgggacaacttaactatacacacggagcaga cgatatggcgtcgctccaccctacacatacaaacacaccagcagagcatatgggctcgctggaacaaca cccgggcataaacggagcagacgatatggcgtcgctccacaaacacaaacatactcagccagcagagc atatgggctcgctggcacctcgcacaaataaacacggagcagacgatatggcgtcgctccaacacacaa ctaacctcacaccagcagagcatatgggctcgctggacacatcgaaacaaataaccggagcagacgat atggcgtcgctccatacaatcaacacacaaataccagcagagcatatgggctcgctggactacacaaatc aacacaacggagcagacgatatggcgtcgctcctacataaacaacacatcaacccagcagagcatatg ggctcgctggaaaccttaacaatctctcagggagcagacgatatggcgtcgctcccatcactcaaataaac accgccagcagagcatatgggctcgctggcacaacacaaacacaacaaaggagcagacgatatggcg tcgctccgcacaactaacaacatcaaaccagcagagcatatgggctcgctggaacatcacccgggcaca acaggagcagacgatatggcgtcgctccatcattcacactacactacaccagcagagcatatgggctcgct ggacaaatacacacaacacaaaggagcagacgatatggcgtcgctccaacatcaacacacaaacccg ccagcagagcatatgggctcgctggcactcaactacacaacactcggagcagacgatatggcgtcgctcc aaacacaatcaaactcaaacccagcagagcatatgggctcgctggcctctcaaatacaagtcgac

First we set out to test if the 24xPP7 array is functional *in vitro*. It was previously shown that PP7ΔFG, which is the aggregation-deficient mutant of PP7, can bind to a single PP7 aptamer *in vitro*[208]. Here we show that the purified PP7ΔFG-mEGFP fusion also binds the single PP7 RNA hairpin *in vitro* and that the fusion protein also needs a free N-terminus (i.e. cleavage of the N-terminal 6xHis tag is required) for binding to the PP7 hairpin. We then showed that PP7ΔFG-mEGFP also binds to the *in vitro* transcribed *clpX*-24xPP7 mRNA (**Figure 3.5**).

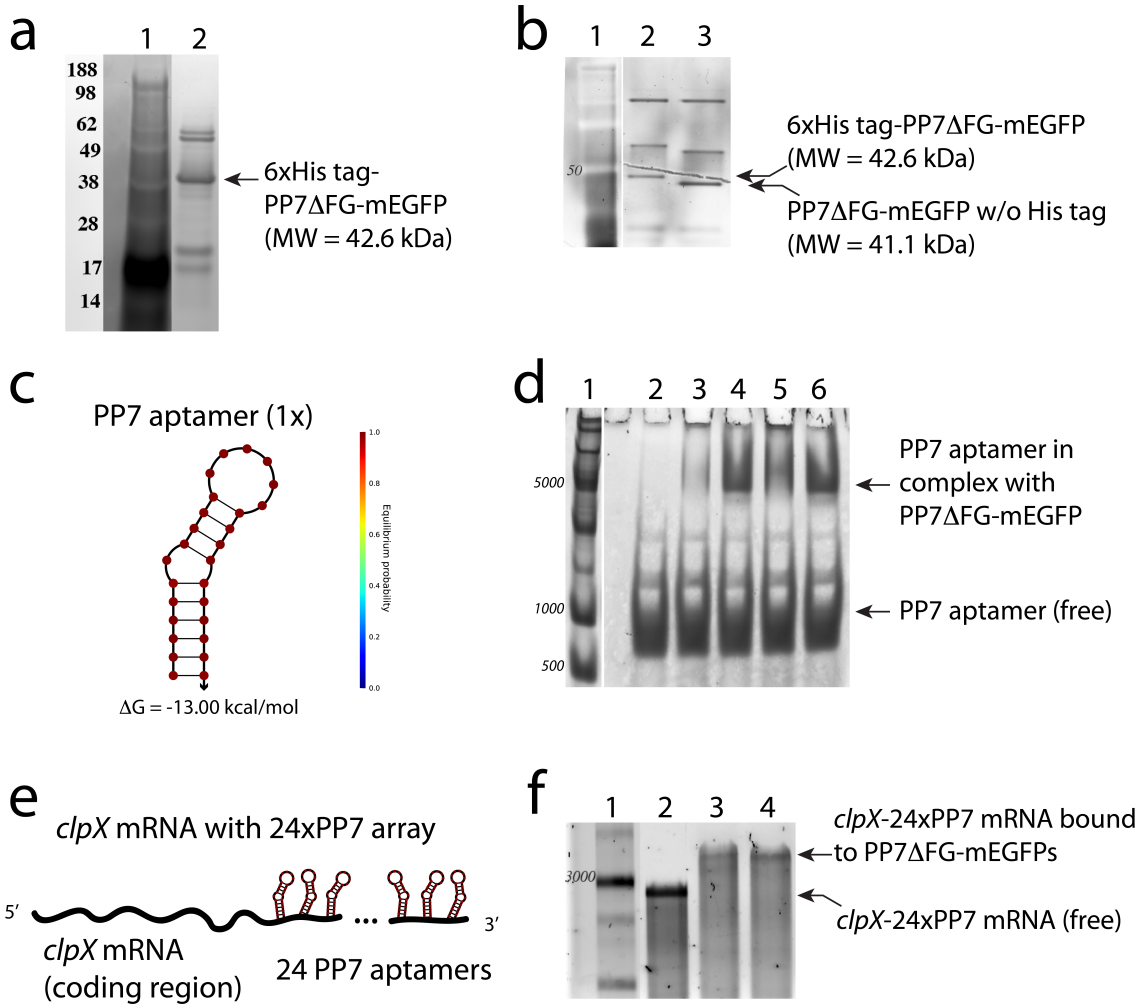


Figure 3.5. PP7ΔFG-mEGFP binds a single PP7 aptamer and the 24xPP7 array *in vitro*. (**a**) The PP7ΔFG-mEGFP fusion protein (lane 2) was purified with an N-terminal 6xHis tag. Protein ladder (lane 1). (**b**) The His tag was removed by thrombin cleavage. Before (lane 2) and after (lane 3) cleavage. Protein ladder (lane 1). (**c**) Predicted secondary structure of the PP7 aptamer. (**d**) Gel shift assay showed that PP7ΔFG-mEGFP binds a single PP7 aptamer. 6xHis-PP7ΔFG -mEGFP (lane 3 and 5) does not bind whereas PP7ΔFG-mEGFP (lane 4 and 6) does bind to the aptamer. Samples in lane 3 and 4 were dialyzed against 1 M NaCl though the high salt treatment had no effect on aptamer binding. Single-stranded (ss) RNA ladder (lane 1) and PP7 aptamer alone (lane 2). (**e**) Schematic of the *clpX* mRNA transcript with the 24xPP7 array. (**f**) Gel shift assay showed that PP7ΔFG-mEGFP binds to the *in vitro* transcribed *clpX*-24xPP7 mRNA (lane 3 and 4). Sample in lane 3 was dialyzed against 1 M NaCl. ssRNA ladder (lane 1) and *clpX*-24xPP7 mRNA alone (lane 2).

The 24xPP7 array was then inserted in the 3' untranslated region (UTR) of the *clpX* mRNA. Live-cell imaging of cells co-expressing the *clpX*-24xPP7 mRNA construct and the PP7ΔFG-mEGFP fusion protein resulted in freely moving fluorescent spots (**Figure 3.6a** and **Video 3.1**). These spots were not observed in a strain that lacked the 24xPP7 array (**Figure 3.6b** and **Video 3.2**). This indicates that the observed spots are not non-specific protein aggregates as occasionally observed for the MS2 system (data not shown) but further experiments will be necessary to show whether each spot corresponds to a single mRNA. The number of spots is reduced 30 min after rifampicin treatment (preliminary experiment, data not shown) further suggesting that the PP7ΔFG-mEGFP molecules form spots because of tethering to the mRNA (and dissociate into monomers after the mRNA is degraded). It was not possible yet to analyze the fluorescence intensities of the PP7 spots because the numbers are too high resulting in many overlapping spots. I will resolve this issue by under-expressing the *clpX*-24xPP7 mRNA.

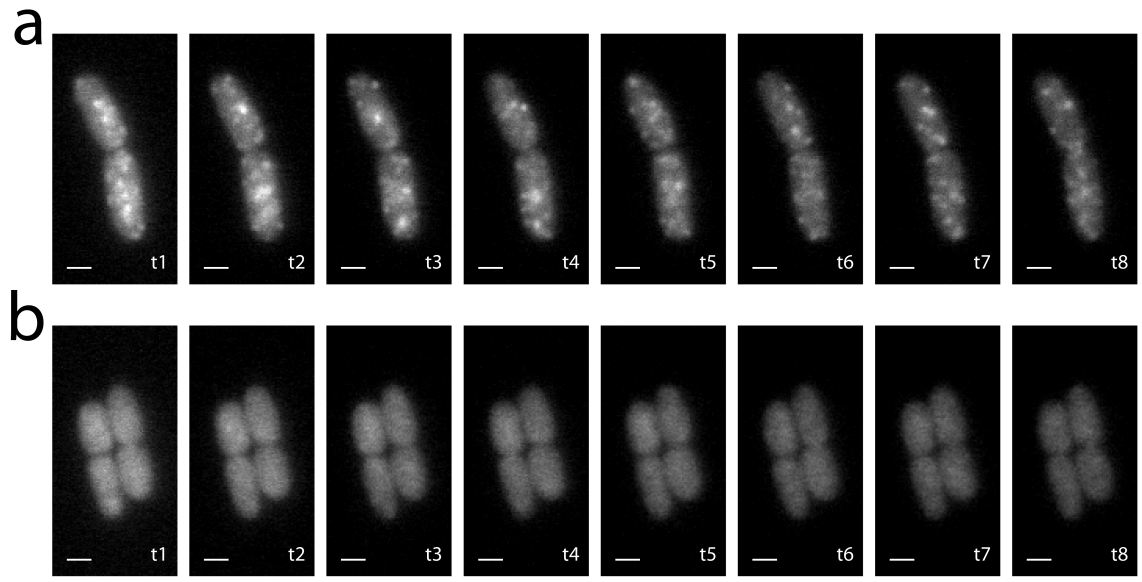


Figure 3.6. PP7ΔFG-mEGFP forms fluorescent spots in live cells that express the *clpX*-24xPP7 mRNA. Fluorescent spots were only detected in the presence of the 24xPP7 array (**a**) GFP time-series of cells with the *clpX*-24xPP7 array (strain DHL1177). (**b**) GFP time-series of cells with untagged *clpX* mRNA (strain DHL1203). Both strains co-express the PP7ΔFG -mEGFP fusion. Scale bars (white) are 1 μm.

3.2.3. Count individual *clpX* mRNA transcripts with FISH and PP7 methods simultaneously

To evaluate the counting assays more quantitatively and to improve parameters of the protocol, we developed a dual-labeling approach to detect the *clpX* mRNA with the two different single-molecule methods described above. Two methods can always be tuned to produce the same average, even if neither one works, and possibly even the same distribution. Requiring that two independent methods agree on individual transcripts is much more rigorous and less subject to experimental artifacts. This assay will ultimately be done in fixed cells with the goal of determining the number of false negatives. If the FISH and PP7 methods indeed can be made to validate each other at a single transcript level, we can further test that the fixation protocol does not introduce artifacts by measuring the same cell population before and after fixation using the PP7 method.

Initial experiments were done *in vitro* with purified components to reduce the complexity. The assay is much more difficult in cells because cell wall digestion, cell membrane permeabilization, probe penetration, inaccessibility of the mRNA targets and fixation issues need to be addressed. Once an *in vitro* method is validated we will therefore move to cell lysates before tackling intact cells.

First we optimized the imaging setup for dual-color single-molecule imaging and confirmed that single surface-immobilized GFP and Quasar570 molecules can be imaged in parallel and that they are spectrally distinct (**Figure 3.7**). This is an absolute requirement for quantitative single-molecule co-localization experiments.

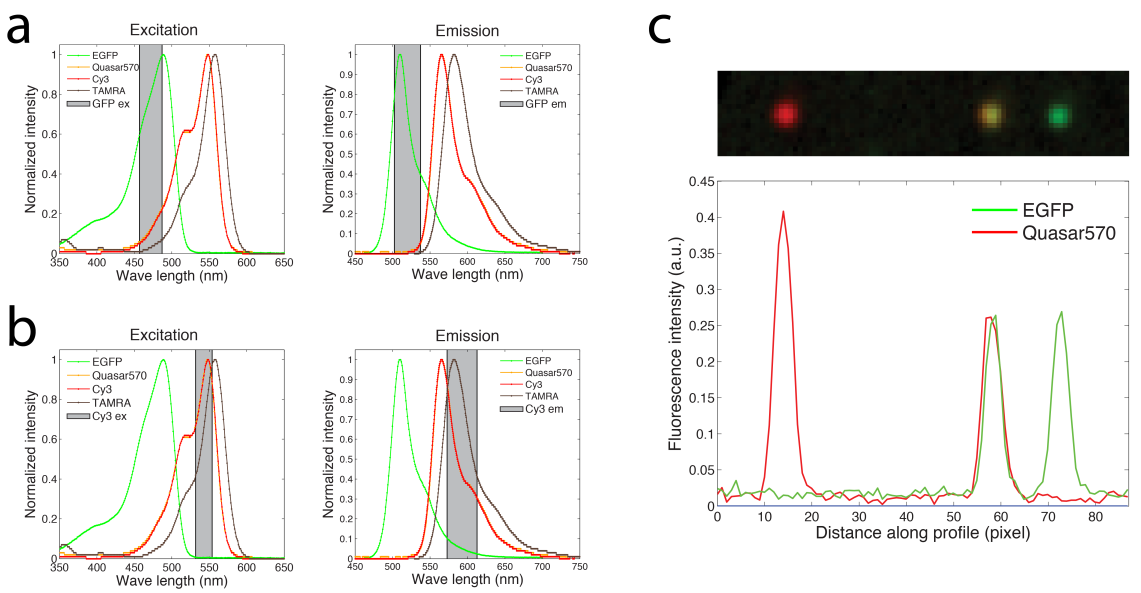


Figure 3.7. GFP and FISH dyes (Cy3, Quasar570 and TAMRA) are spectrally distinct. The excitation and emission spectra of the Cy3 (red) and Quasar570 (orange) dyes are virtually undistinguishable but TAMRA (brown) is more red-shifted. (**a**) GFP filter set with bandpass filter (gray). (**b**) Cy3 filter set with bandpass filter (gray). The raw data of the absorbance and emission spectra of the dyes were obtained from Biosearch Technologies, Inc. and of EGFP from Roger Tsien's homepage (http://www.tsienlab.ucsd.edu/Documents.htm). (**c**) Intensity profile of single EGFP (green) and Quasar570 (red) molecules. Left peak corresponds to a Quasar570 molecule, right peak to an EGFP molecule and middle peak to a Quasar570 molecule that is co-localized with an EGFP molecule.

Single surface-immobilized GFP molecules were imaged with epi-fluorescence microscopy to determine the average GFP fluorescence intensity of a single molecule. It has been previously demonstrated that single surface-immobilized GFP molecules can be detected with standard epi-fluorescence microscopy and a normal CCD camera[215]. Here, the average fluorescence of a single GFP molecule was determined by two different methods: analysis of fluorescence intensities (**Figure 3.8a** and **b**) and photobleaching steps (**Figure 3.8c** and **d**). The average fluorescent intensity of a single GFP molecule corresponds to ~120 fluorescence counts under our imaging conditions.

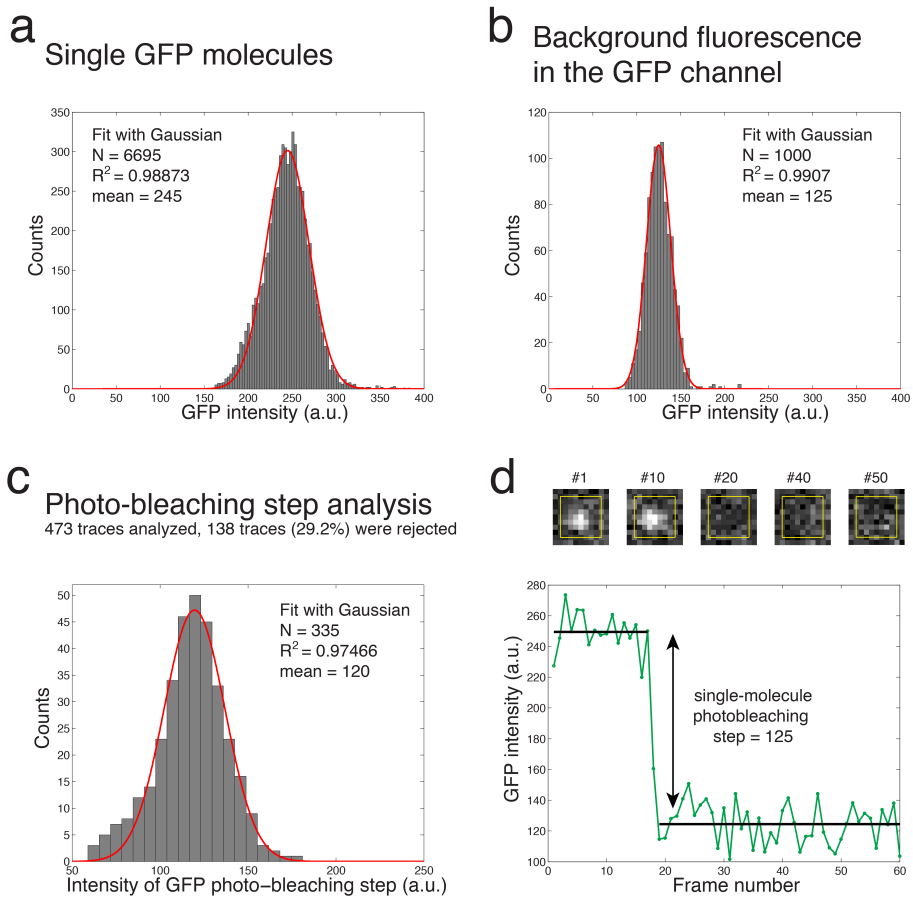


Figure 3.8. The average fluorescent intensity of a GFP molecule corresponds to ~120 fluorescence counts as determined by the analysis of fluorescence intensities (**a**) and photobleaching steps (**b**). The GFP molecule used has the AP tag[216] (hence is biotinyated), for surface attachment, engineered in a surface exposed loop of mEGFP[126]. It is also fused to the SNAPf tag but not labeled with a dye. This GFP fusion was named iAPmEGFP-SNAPf (purified protein courtesy of Burak Okumus).

Finally, individual surface-immobilized *clpX*-24xPP7 mRNA transcripts were simultaneously labeled with *clpX* Quasar570 FISH probes and PP7ΔFG-mEGFP molecules. Single-molecule imaging revealed three subpopulations (**Figure 3.9a**).

Figure 3.9

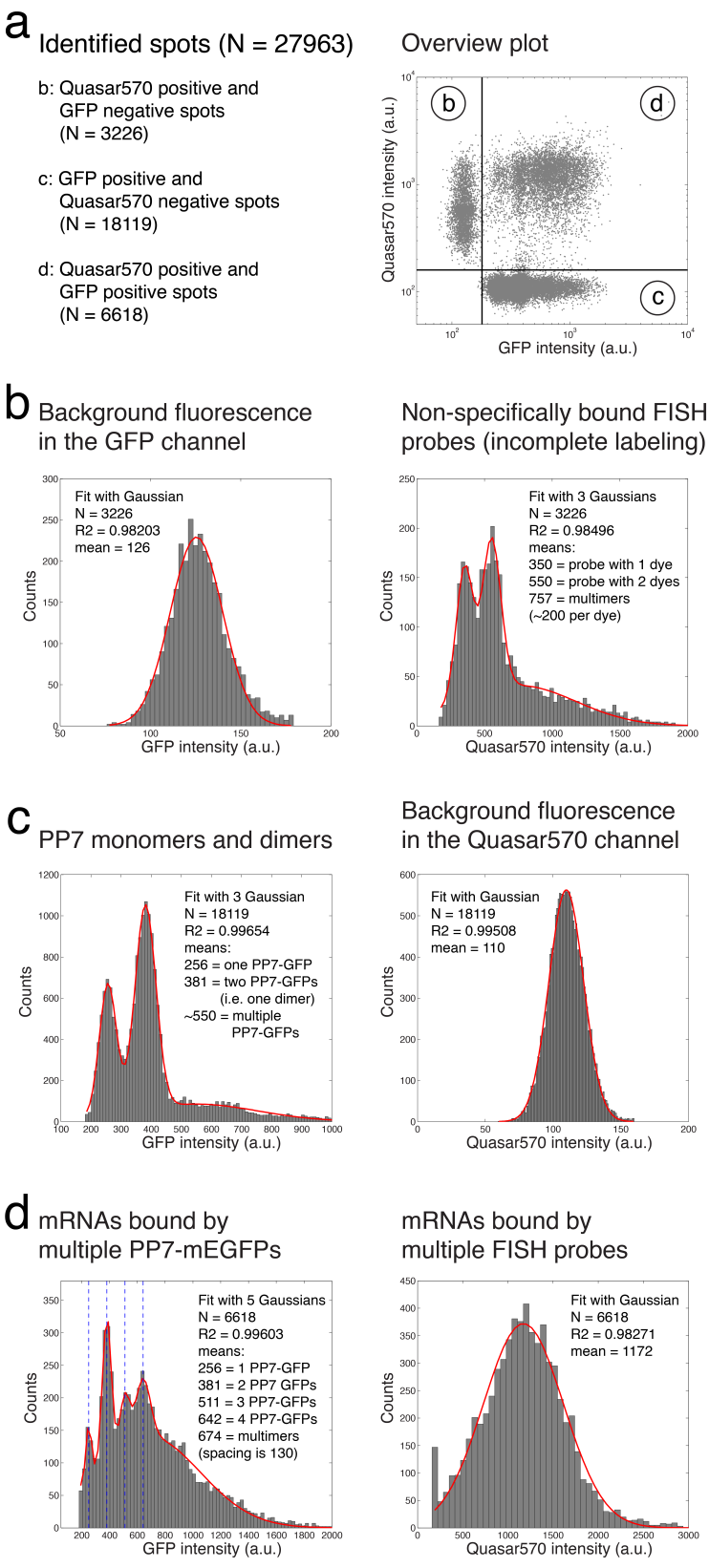

a
Identified spots (N = 27963)
b: Quasar570 positive and GFP negative spots (N = 3226)
c: GFP positive and Quasar570 negative spots (N = 18119)
d: Quasar570 positive and GFP positive spots (N = 6618)
Overview plot
Quasar570 intensity (a.u.)
GFP intensity (a.u.)
b
Background fluorescence in the GFP channel
Fit with Gaussian
N = 3226
R2 = 0.98203
mean = 126
Counts
GFP intensity (a.u.)
Non-specifically bound FISH probes (incomplete labeling)
Fit with 3 Gaussians
N = 3226
R2 = 0.98496
means:
350 = probe with 1 dye
550 = probe with 2 dyes
757 = multimers
(~200 per dye)
Counts
Quasar570 intensity (a.u.)
c
PP7 monomers and dimers
Fit with 3 Gaussian
N = 18119
R2 = 0.99654
means:
256 = one PP7-GFP
381 = two PP7-GFPs
(i.e. one dimer)
~550 = multiple PP7-GFPs
Counts
GFP intensity (a.u.)
Background fluorescence in the Quasar570 channel
Fit with Gaussian
N = 18119
R2 = 0.99508
mean = 110
Counts
Quasar570 intensity (a.u.)
d
mRNAs bound by multiple PP7-mEGFPs
Fit with 5 Gaussians
N = 6618
R2 = 0.99603
means:
256 = 1 PP7-GFP
381 = 2 PP7 GFPs
511 = 3 PP7-GFPs
642 = 4 PP7-GFPs
674 = multimers
(spacing is 130)
Counts
GFP intensity (a.u.)
mRNAs bound by multiple FISH probes
Fit with Gaussian
N = 6618
R2 = 0.98271
mean = 1172
Counts
Quasar570 intensity (a.u.)

Figure 3.9 (Continued).

Figure 3.9. Simultaneous detection of *clpX*-24xPP7 mRNA transcripts with the *clpX* Quasar570 FISH probes and the PP7ΔFG-mEGFP protein. (**a**) Data from multiple fields was pooled and 27963 spots were computationally identified (see Materials and Methods). The GFP and Quasar570 intensities of every spot were quantified. 3226 spots were Quasar570 positive and GFP negative, 18119 spots were GFP positive and Quasar570 negative and 6618 spots were positive for both, GFP and Quasar570. (**b**) The two Quasar570 peaks correspond to the FISH probes with one dye (mean1 = 350) and two dyes (mean2 = 550). (**c**) The GFP peaks correspond to non-specifically bound PP7ΔFG-mEGFP monomers (mean1 = 256), dimers (mean2 = 381), and higher-order aggregates (mean3 = 550). (**d**) Spots with co-localized GFP and Quasar570 signal likely correspond to individual mRNAs. The histogram of GFP intensities shows multiple peaks and the spacing between the peaks is ~130 units, which corresponds approximately to the average intensity of a single GFP molecule (**Figure 3.8**). The histogram of Quasar570 intensities corresponds to multiple FISH probes bound to the *clpX*-24xPP7 mRNA. The peak is broad likely because of incomplete labeling of the FISH probes and incomplete binding of the probes to the mRNA.

The first subpopulation of spots (i.e. Quasar570 positive and GFP negative) corresponds to non-specifically bound FISH probes. The two large peaks of the histogram again show that the *clpX* Quasar570 FISH probes are incompletely labeled as previously observed and discussed (**Figure 3.3c** and **d**). The second subpopulation (i.e. GFP positive and Quasar570 negative) corresponds to PP7 monomers and dimers. These molecules are probably non-specifically bound to the imaging surface (or aggregates) since they do not overlap with the FISH probes. The third subpopulation (i.e. Quasar570 positive and GFP positive) likely corresponds to real mRNAs since each spot has a Quasar570 and GFP signal above background fluorescence.

The histogram of GFP fluorescence intensities shows multiple peaks and Gaussian fitting indicates that the peaks correspond to integer numbers of PP7-mEGFP monomers (not dimers) that are bound to the 24xPP7 array of the *clpX* mRNA. The number of bound GFP molecules is surprisingly low and far from complete. Potential explanations are incomplete folding of the PP7 hairpins or that excessive washing after the incubation greatly reduced the number of PP7 molecules. One 24xPP7 array could theoretically be bound by 48 GFP molecules since each hairpin can accommodate one PP7 dimer (i.e. 2 PP7ΔFG-GFP molecules). It is believed that PP7 has to dimerize before it can bind to the PP7 hairpin though this would suggest that one should obtain a histogram with peaks that are spaced by a distance that is equal to two GFP molecules and not one as we observed here. It is unlikely that this is because of incomplete maturation of the GFPs since the GFP proteins were purified from *E. coli*, dialyzed overnight and then stored at –80 ˚C for at least a few days before they were used; hence ample time to complete chromophore maturation.

A recent study showed that a single-chain tandem dimer of PP7ΔFG-GFP is more efficient in aptamer binding since linking the subunits eliminates the intermolecular dimerization process[145]. It will be interesting to use this tandem dimer for the *in vitro clpX*

FISH experiments and also for the *in vivo* experiments. Also, the surface preparation was not optimal for this type of experiment since BSA-coated imaging surfaces display non-specific protein sticking. The use of a polyethylene glycol (PEG)-coated imaging surface would be more appropriate to prevent non-specific protein aggregation on the surface (Burak Okumus, personal communication) and will be tried in the future.

The histogram of Quasar570 intensities corresponds to multiple bound *clpX* FISH probes but the peak is rather broad. This is partly because the FISH probes are not completely labeled but the data also suggests that the probe binding to the mRNA is not complete and exhibits large heterogeneity under the incubation conditions used. The *in vitro* assay, presented here, provides detailed information about the binding of the *clpX* FISH probes and the PP7 protein. Existing problems, which would be otherwise very hard to detect, were revealed. This is very valuable information for optimizing the parameters before carrying out the dual-labeling experiments in cells.

3.3. Discussion

I presented in this chapter approaches to develop accurate mRNA counting methods in *E. coli*, motivated by the absence of accuracy controls in previous studies, and the many reasons why such methods in fact may be non-quantitative. Through carefully following established procedures, we indeed demonstrated many potential issues with both single-molecule FISH and assays based on the RNA-binding protein PP7. However, both the assays developed and the problems identified will also help to improve the method. For example, the FISH probe evaluation of two different sets of multi-labeled *clpX* probe sets revealed that incomplete labeling is a great challenge since it creates great heterogeneity in the intensity per molecule and thereby renders quantification difficult to impossible. However, it should be feasible to improve probe labeling and purify the completely labeled probes from the incomplete ones. The *in vitro* analysis of individual FISH probe spot intensities and photobleaching steps can be used to accurately determine such labeling problems and will be used to identify a quantitative probe set with a narrow unimodal intensity distribution. If completely labeled probes cannot be obtained, the alternative method[196] that uses a large number of singly labeled probes will be used instead. However, such probes may simply hide the problem since any partial labeling makes the molecules invisible altogether.

After fully labeled probes are obtained, FISH against the *in vitro* transcribed *clpX* mRNA can be used to determine the heterogeneity in intensity per mRNA transcript, and to optimize the choice of probes. A narrow intensity distribution per single transcript is important to be able to infer the number of mRNAs contained within any multi-transcript foci.

The PP7 mRNA imaging reporter served two roles here. First it was employed as an important tool for establishing the cross-validation test and for characterization of the FISH assay. Second, PP7 is valuable as a reporter system for mRNA imaging in live bacteria and will likely replace MS2 since the PP7 system is more quantitative and does not suffer from mRNA mislocalization and artifactual foci formation problems. It will therefore be interesting to test whether the tdPP7ΔFG-mEGFP construct increases the signal-to-noise ratio and lead to higher occupancy of the 24xPP7 array. It also needs to be tested if the 24xPP7 array negatively impacts mRNA degradation and whether all PP7 spots that are observed in the live-cell movies have mRNA transcripts bound (i.e. are not false positives). If the 24xPP7 array interferes with mRNA degradation, it could be inserted in the 5' UTR though interference with degradation will not matter for the cross-validation experiments as long as the mRNA concentration will not be too high.

The ultimate experiment is to combine both reporter systems, FISH probes and the PP7 protein, and evaluate FISH measurements in cells. This will allow minimizing the fraction of false-negatives and establishing reliable protocols for truly single-molecule accuracy measurements of mRNA transcripts in *E. coli*.

3.4. Materials and Methods

Plasmid constructions

All plasmids used in this study are listed in **Table 3.3**.

Table 3.3

Plasmid	Description	Antibiotic marker	Reference
pDHL580	pUC19-linker-mGFPmut3-FRT Kan FRT	Amp, Kan	[205]
pDHL782	pSC101-P_{LlacO1}-weak RBS-mCherry ssrA tag-T1 terminator-*lacI*^{q}	Amp	Lab collection
pDHL962	pUC19-linker-mEGFP-FRT Kan FRT	Amp, Kan	Lab collection
pDHL1002	pUC19-linker-Spinach-FRT Kan FRT	Amp, Kan	This study
pDHL1091	pSC101-P_{LlacO1}- PP7ΔFG-mEGFP-T1 terminator	Amp	This study
pDHL1134	pUC57-24xPP7 array	Kan	This study
pDHL1138	pET28a-*clpX*-T1 terminator	Kan	This study
pDHL1149	pSC101-P_{recC}-PP7ΔFG-mEGFP-T1 terminator	Amp	This study
pDHL1152	pUC19-linker-24xPP7 array-FRT Kan FRT	Amp, Kan	This study
pDHL1162	pSC101-P_{LlacO1}-weak RBS-PP7ΔFG-mEGFP-T1 terminator	Amp	This study
pDHL1165	pET28a-PP7ΔFG-mEGFP-T1 terminator	Kan	This study
pDHL1194	pET28a-*clpX*-24xPP7 array-T1 terminator	Kan	This study
pPM16	pSC101-P_{LlacO1}-Venus-T1 terminator	Amp	[205]
pET28b-Spinach	pET28b-$tRNA_{Lys}$ 5' stem-Spinach aptamer-$tRNA_{Lys}$ 3' stem-T7 terminator	Kan	[210]

The construction of the plasmids was verified by analytical restriction digests. All cloning steps that involved PCR amplification were validated by DNA sequencing. AccuPrime Pfx DNA polymerase (Invitrogen) was used for PCR amplification. The restriction enzymes were purchased from NEB and used according to the manufacturer's instructions.

Plasmid pDHL1002 was built by amplifying the Spinach aptamer with the tRNA scaffold from pET28b-Spinach with primers DHL_P577_F and DHL_P578_R. The PCR product

was gel purified and ligated into pDHL580, which was cut with SacI and XmaI, by using isothermal assembly[181].

Plasmid pDHL1091 was built by using two inserts. PP7ΔFG was amplified from a PP7-containg plasmid[217] (courteously of Camille Delebecque) with primers DHL_P676_F and DHL_P677_R and mEGFP[126] (EGFP A206K[31]) was amplified from pDHL962 with primers DHL_P678_F and DHL_P679_R. The mEGFP part also includes a short linker (SGGGGSK). The PP7 used corresponds to the aggregation-deficient variant PP7ΔFG, which has the FG loop (C68-E75) of the wild-type PP7 replaced against a serine (S68) and a glycine (G69) residue[208]. The PCR products were gel purified and inserted into pPM16, which was digested BamHI and HindIII, by using isothermal assembly[181].

pDHL1134 was obtained by DNA synthesis (Genewiz). The 24xPP7 array contains 24x binding sites for the RNA-binding protein PP7. Two different PP7 binding sites, PP7_s1 (ggagcgacgccatatcgtctgctcc) and PP7_s2 (ccagcagagcatatgggctcgctgg), were alternated and the individual PP7 binding sites are separated from each other by various 20 bp inert spacer sequences. The spacer sequences were optimized to not interfere with folding of the PP7 aptamers with the software package NUPACK (http://www.nupack.org/). The DNA sequence of 24xPP7 array is listed in **Table 3.2**.

pDHL1138 was built by amplifying the *clpX* coding region from genomic DNA (made from the strain MC4100) with primers DHL_P686_F and DHL_P687_R. The PCR product was gel purified and ligated into pET28a (Novagen), which was previously digested by NdeI and BlpI, with isothermal assembly[181].

pDHL1149 was built by PCR amplifying the *recC* promoter from genomic DNA (made from the strain MC4100) with primers DHL_P720_F and DHL_P721_R. The *recC* promoter PCR product was gel purified and ligated into the XhoI/BamHI-digested pDHL1091 plasmid with isothermal assembly[181].

pDHL1152 was built by PCR amplifying the 24xPP7 array from plasmid pDHL1134 with primers DHL_P718_F and DHL_P719_R. The PCR product was gel purified and ligated into pDHL580, which was digested with SacI and XmaI, with isothermal assembly[181].

pDHL1162 was built by PCR amplifying the P_{LlacO1} promoter[176] including a weak RBS (ATTTCACACAGGAAACC) from plasmid pDHL820 with primers DHL_P722_F and DHL_P723_R. The resulting PCR product was gel purified and ligated into the XhoI/BamHI-digested pDHL1091 plasmid with isothermal assembly[181]. The resulting plasmid produces high levels of mRNA but inefficient translation, due to the weak RBS, leading to low protein levels. The variability in protein levels is low within an isogenic population of cells (data not shown) because inefficient translation averages out upstream mRNA fluctuations[218].

pDHL1165 was built by amplifying the region corresponding to PP7ΔFG-linker-mEGFP with primers DHL_P729_F and DHL_P730_R from plasmid pDHL1091. The PCR product was gel purified and ligated into the NdeI/BlpI-digested pET28a vector with isothermal assembly[181].

pDHL1194 was built by PCR amplifying the 24xPP7 array from plasmid pDHL1138. The 24xPP7 array PCR product was gel purified and inserted into the pDHL1138 vector, which was cut with BlpI, by using isothermal assembly[181].

Primers

All primers were purchased from Integrated DNA Technologies, Inc. (IDT) and are listed in **Table 3.4**.

Table 3.4

Primer	Sequence
DHL_P171_R	ggagataaaatccccccttttttggttaactaattgtatgggaatggttaa-attccggggatccgtcgacc
DHL_P173_F	ctgagtttattggtcgtctgccgg
DHL_P556_R	atgacacgactgtgcttcacgc
DHL_P577_F	cggccagtgaattcgagctctaacgctgacgaccgactgcccggatagctcagtcggtag
DHL_P578_R	aaagtataggaacttcgaagcagctccagcctacaccccgggtggcgcccgaacagggac
DHL_P582_R	aaccgttgctgatttatggcaagccggaagcgcaacaggcatctggtgaa-taacgctgacgaccgactgcc
DHL_P676_F	cactgaccgaattcattaaagaggagaaaggatccatgtccaaaaccatcgttctttcgg
DHL_P677_R	cctcgcccttgctcactttactgccaccgccaccgctacggcccagcggcacaag
DHL_P678_F	aagatcttgtcgtcaaccttgtgccgctgggccgtagcggtggcggtggcagtaa
DHL_P679_R	ttatttgatgcctctagactcagctaattaagcttacttgtacagctcgtccatgccgag
DHL_P686_F	gcagcggcctggtgccgcgcggcagccatatgacagataaacgcaaagatggctc
DHL_P687_R	gccccaaggggttatgctagttattgctcagcttattcaccagatgcctgttgcg
DHL_P718_F	ggccagtgaattcgagctctaacgctgacgaccgactgcctgcagcaaacaaactacccg
DHL_P719_R	aacttcgaagcagctccagcctacaccccgggcccgtcgacttgtatttgagagg
DHL_P720_F	atagggcgaattgggtaccgggccccccctcgagatttcgccacctgccaactgg
DHL_P721_R	cgaaagaacgatggttttggacatggatccagcggctcctgactactgacgattc
DHL_P722_F	tagggcgaattgggtaccgggccccccctcgagaattgtgagcggataacaattg
DHL_P723_R	cgccgaccgaaagaacgatggttttggacatggatccggtttcctgtgtgaaatg
DHL_P727_F	ggaagcgcaacaggcatctggtgaataagctaacgctgacgaccgactgcc
DHL_P728_R	gaggccccaaggggttatgctagttattgccccgtcgacttgtatttgagaggc
DHL_P729_F	agcagcggcctggtgccgcgcggcagccatatgtccaaaaccatcgttctttcgg
DHL_P730_R	cccaaggggttatgctagttattgctcagcttacttgtacagctcgtccatgccg

E. coli strains

All *E. coli* strains used in this study are listed in the **Table 3.5**.

Table 3.5

Strain	Description	Antibiotic marker	Reference
MC4100	Wild-type (wt) *E. coli* strain	–	Lab collection
DH5α	Standard cloning strain	–	Lab collection
NEB Turbo	Standard cloning strain	–	Lab collection
NEB10β	Standard cloning strain	–	Lab collection
C41	BL21(DE3) derivative, higher tolerance to over-expression of toxic proteins	–	Lab collection
DHL708	MC4100 Δ*clpPX*	–	[205]
DHL749	MC4100 ara+ pKD46	Amp	This study

Table 3.5 (Continued).

DHL1002	NEB Turbo pDHL1002	Amp, Kan	This study
DHL1006	MC4100 pDHL1001	Amp	This study
DHL1011	MC4100 ara+ *clpX*-Spinach-FRT Kan FRT	Kan	This study
DHL1091	NEB Turbo pDHL1091	Amp	This study
DHL1092	MC4100 pDHL1091	Amp	This study
DHL1134	NEB10β pDHL1134	Kan	This study
DHL1149	DH5α pDHL1149	Amp	This study
DHL1150	MC4100 *clpX*-12xPP7 array	–	This study
DHL1152	NEB10β pDHL1152	Amp, Kan	This study
DHL1158	DHL749 *clpX*-24xPP7 array-FRT Kan FRT	Amp, Kan	This study
DHL1162	NEB5α pDHL1162	Amp	This study
DHL1165	NEB5α pDHL1165	Kan	This study
DHL1168	C41 pDHL1165	Kan	This study
DHL1174	MC4100 ara+ *clpX*-24xPP7 array	–	This study
DHL1194	NEB10β pDHL1194	Kan	This study
DHL1203	MC4100 pDHL1162	Amp	This study

Strain DHL1011 was built by amplifying the Spinach aptamer and the FRT-flanked Kan marker from plasmid pDHL1002 with primers DHL_P582_F and DHL_P171_R. The PCR product was DpnI-digested, purified and then transformed into DHL749 and integrated into the chromosomal right downstream of the *clpX* stop codon. The lambda Red-mediated recombination[172] used, was previously described in great depth[205]. The pKD46 plasmid was eliminated from strain DHL1011 but the *clpX*-Spinach-FRT Kan FRT allele was not transferred into a fresh MC4100 background by P1 transduction. Integration was confirmed by colony PCR and the integration site was sequenced (Genewiz) with primers DHL_P173_F and DHL_P556_R.

Strain DHL1158 was built by PCR amplifying the 24xPP7 array and the FRT-Kan cassette from pDHL1152 with primers DHL_P582_F and DHL_P171_R. The construct was integrated into the chromosome right downstream of the *clpX* stop codon (as described fro strain DHL1011). Integration was confirmed by colony PCR. The pKD46 plasmid was cured from positive colonies by growth at 42 ˚C. The integration site was sequenced (Genewiz) with primers DHL_P173_F and DHL_P556_R. Strain DHL1158

was transformed with pCP20 to excise the Kan marker from the chromosome. Loss of the Kan marker was confirmed by PCR and sensitivity to Kan. The pCP20 plasmid was by growth at 42 ˚C. The resulting strain was named DHL1174.

FISH probes and anchor oligos

The 50-nt long FISH probes were designed by following previously described guide lines[212]. A custom-written MATLAB script was used to find all 50-nt long DNA segments that are reverse complements (essentially reverse primers) to the coding region of the *clpX* gene and have a GC-content of 50%. This created a pool of 135 FISH probe candidates and all probes were further evaluated. Chemical properties, like melting temperature (Tm), Gibb's free energy of hairpin formation ($\Delta G_{hairpin}$) and Gibb's free energy of self-dimerization ($\Delta G_{self\text{-}dimer}$) were calculated with OligoAnalyzer 3.1 (http://www.idtdna.com/analyzer/applications/oligoanalyzer/) or NUPACK (www.nupack.org) for all potential FISH probes. The parameters used for the secondary structure calculations were 1 μM DNA oligo, 100 mM NaCl and a temperature of 37 ˚C. The dimer formation was calculated with NUPACK by allowing a maximum complex size of 2. FISH probes that were prone to hairpin formation (i.e. $\Delta G_{hairpin}$ more negative than –3.0 kcal/mol based on OligoAnalyzer 3.1 or –5.5 kcal/mol based on the NUPACK calculation) or dimer formation were eliminated from the pool of potential FISH probes. Each FISH probe was then searched against the *E. coli* genome by using the BLAST tool of the ECOCYC webpage (http://ecocyc.org/ECOLI/blast.html) and probes with specific off-target binding sites were rejected.

The best five remaining FISH probes (i.e. *clpX*-77, *clpX*-433, *clpX*-674, *clpX*-853 and *clpX*-999, see **Table 3.6**) were synthesized and three Quasar 570 dyes were co-synthetically incorporated by using Quasar 570-coupled thymidine phosphoramidites

(Biosearch Technologies). For the Cy3-labeled *clpX* FISH probes, only four probes were used (i.e. *clpX*-77, *clpX*-433, *clpX*-853 and *clpX*-999). The probes were synthesized and labeled by the Singer laboratory (Albert Einstein College of Medicine) with five Cy3 dyes (assuming 100% labeling efficiency) using NHS chemistry[212,219].

Table 3.6

Probe	Sequence	Comment
clpX-77	aTcaacacattcgtcgcagataTacacggatggaccggcaatcagctTgc OR aTcaacacatTcgtcgcagaTatacacggaTggaccggcaatcagcTtgc	[*] [**]
clpX-433	aTgttttcaacgtcttcacccacaTaaccggcttcggtcagtgtagTcgc OR aTgttttcaacgTcttcacccacaTaaccggcttcggTcagtgtagTcgc	[*] [**]
clpX-674	agaggTatcaacctgcaagaattccTgctgcggatgtttacgcccaccTt	[*]
clpX-853	aTaagaccaaacttgatcagatctTccggttcaacctgcgccagcagcTc OR aTaagaccaaacTtgatcagatctTccggttcaaccTgcgccagcagcTc	[*] [**]
clpX-999	aTtccagatccacgccttccagatTaaacagcgcctgatactgcttggTc OR atTccagatccacgccTtccagatTaaacagcgcctgaTactgcttggTc	[*] [**]

[*] The capitalized thymidine residues (i.e. T) served as dye attachment sites for the Quasar570 dyes. [**] The capitalized thymidine residues (i.e. T) served as dye attachment sites for the Cy3 dyes.

The biotinylated anchor probes for the *clpX* mRNA and the PP7 RNA oligos (**Table 3.7**) were ordered from IDT DNA and were resuspended in RNase-free water.

Table 3.7

Probe	Sequence	Comment
DHL_P408	Biotin-gcaagctgattgccggtccatccgtgtatatctgcgacgaatgtgttgat	-
BO_P17	Alexa488- atcaacacattcgtcgcagatatacacggatggaccggcaatcagcttgctttttttttttttttttttt –Biotin	-
DHL_P746	gctgctgtgatgatgatgatgatggctgctgcccatggtatatctccttc ttttttttttttttttttt-Biotin	-
DHL_P747	gcgcacagaagauauggcuucgugcgc	RNA oligo
DHL_P759	Biotin-uuuuuuuuuu-gcgcacagaagauauggcuucgugcgc	RNA oligo

Estimation of the labeling efficiency of the Cy3-labeled *clpX* FISH probes and the *clpX*-77 Quasar570 FISH probe

1 ng of Cy3-labeled *clpX* FISH probes was diluted in 1 ml ddH_2O and pipetted on a poly-lysine-coated coverslip, which was then incubated for 30 min in the dark. The coverslip was washed 3x with 1 ml ddH_2O with 5 min incubations between washes. The coverslip was air dried and attached to a microscope glass slide with double-sided sticky tape to create a flow chamber. The flow chamber was filled with ddH_2O, sealed with 5-min epoxy and then imaged (see below).

10 µM *clpX*-77 FISH probe (Quasar570) was mixed with 10 µM DHL_P408, which is complementary to the FISH probe and has a 5' biotin for surface attachment, and equilibrated in buffer A (10 mM Tis (pH 8.0) and 100 mM NaCl). The solution was then incubated for 2 min in a 95 ˚C water bath followed by > 30 min incubation at room temperature. The annealed double-stranded FISH probe was then diluted to 10 pM in buffer B (5 mM Tris (pH 8.0), 1 mM EDTA and 1 mM $MgCl_2$) with 1 mg/ml BSA (Sigma-Aldrich) and surface-immobilized with BSA-biotin and streptavidin chemistry in a flow-chamber (see below). The flow chamber volume was exchanged against Gloxy buffer[196] to reduce photobleaching. The FISH probes were imaged with an epi-fluorescence microscope (see below).

The acquired images were loaded into Matlab (MathWorks), background subtracted, bandpass filtered with the bpass function and local maxima, corresponding to individual molecules, were localized with the pkfnd function (http://physics.georgetown.edu/matlab/code.html). The fluorescence intensity per spot was determined by calculated average fluorescence intensity within a 7 x 7 box around the molecule and histograms of the fluorescence intensity were generated. The histogram of the *clpX* Cy3 FISH probe spots was fitted to the sum of four Gaussians:

$y = a_1\, e^{-\frac{1}{2}\left(\frac{x-b_1}{c_1}\right)^2} + a_2\, e^{-\frac{1}{2}\left(\frac{x-b_2}{c_2}\right)^2} + a_3\, e^{-\frac{1}{2}\left(\frac{x-b_3}{c_3}\right)^2} + a_4\, e^{-\frac{1}{2}\left(\frac{x-b_4}{c_4}\right)^2}$ where $b_n = n\, b_1$ and $c_n = n\, c_1$ for $n = 1, 2, 3$ and 4. The equation has 6 free fitting parameters (i.e. four scaling parameters $a_1 - a_4$, mean b_1 and standard deviation c_1).

For the *clpX*-77 Quasar 570 FISH probes, the histogram was fitted to the sum of three Gaussians: $y = a_1\, e^{-\frac{1}{2}\left(\frac{x-b_1}{c_1}\right)^2} + a_2\, e^{-\frac{1}{2}\left(\frac{x-b_2}{c_2}\right)^2} + a_3\, e^{-\frac{1}{2}\left(\frac{x-b_3}{c_3}\right)^2}$ where $b_n = n\, b_1$ and $c_n = n\, c_1$ for $n = 1, 2$ and 3. This equation has 5 free fitting parameters (i.e. $a_1 - a_3$, b_1 and c_1). Fitting was done with Matlab using the fminsearch function to minimize the sum of the squared vertical residues. The individual areas under the three Gaussian curves were calculated with $A_n = a_n\, c_1 \sqrt{2\pi}$ where A_n is the area under the n^{th} Gaussian. Fitting of the third Gaussian (of *clpX*-77 Quasar 570 FISH probe histogram) could be improved by making c_3 a free fitting parameter (i.e. 6 free parameters in total), which resulted in $c_3 = 2.31\, c_1$ (instead of $c_3 = 3\, c_1$).

Purification of the PP7ΔFG-mEGFP protein

The PP7ΔFG-mEGFP protein was purified from *E. coli* strain DHL1168 by following a standard protocol for the purification of His-tagged proteins. The protein concentration was determined with the Bio-Rad DC Protein Assay (BioRad) and with the NanoDrop spectrophotometer (ThermoScientific). The protein was then dialyzed overnight against 1x PBS 10% (v/v) glycerol. The N-terminal 6xHis tag was removed by digestion with thrombin (GE Healthcare Life Sciences, cat# 27-0846-01). The protein was then aliquoted, frozen in liquid nitrogen and stored at –80˚C. Fresh aliquots, which were thawed on ice, were used for the experiments.

In vitro transcription

The *clpX*-24xPP7 mRNA was transcribed from a linear double-stranded DNA template (pDHL1194 digested with SalI) with the AmpliscribeT7 Flash kit (Epicentre Biotechnologies) overnight at 37 ˚C following the manufacturer's instructions. The reaction mix was digested with RNase-free DNase (10,000 U/mg) for 30 min at 37 ˚C to eliminate the DNA template of the *in vitro* transcription. The resulting full-length mRNA was ethanol purified and resuspended in DEPC-treated water. The RNA concentration was measured with a NanoDrop spectrophotometer (ThermoScientific).

Folding of the PP7 aptamer and 24xPP7 array

The PP7 hairpin-containing RNAs (i.e. single PP7 aptamer or the *clpX*-24xPP7 mRNA) were folded by incubating 1 µM of the respective RNA at 95 ºC for 1 min followed by snap cooling of the RNA on ice for 30 min. The folding buffer contained 25 mM HEPES (pH 7.0), 25 mM NaCl, 1 mM EDTA and RNase inhibitor.

Gel shift experiment

The gel shift experiments were performed by adapting a previously described protocol[208] that was developed for analyzing the binding of PP7ΔFG to a PP7 aptamer. 100 nM of the PP7 RNA hairpin (RNA oligo DHL747 or DHL759) or 100 nM of *in vitro* transcribed *clpX*-24xPP7 mRNA were mixed with several serial dilutions of purified PP7ΔFG-mEGFP protein (0.8–2 µM) and then equilibrated in buffer containing 10 mM Tris (pH 7.5), 25 mM NaCl, 0.1 mM EDTA, 0.01 mg/ml yeast tRNA (Ambion), 50 µg/ml heparin and 0.01% (v/v) IGEPAL CA630 for 1–2 h at room temperature. The RNA-protein complexes were then separated from free RNA by native polyacrylamide gel electrophoresis (6% (w/v) acrylamide, 0.5x TBE) at room temperature. The gels were

run at 120 V for 40 min with 0.5x TBE running buffer. The bands were visualized by using a SYBR safe stain and imaged using a Typhoon 9400 scanner.

Binding of PP7-mEGFP to the *in vitro* transcribed *clpX*-24xPP7 mRNA and hybridization with FISH probes

First, 100 nM of *clpX*-24xPP7 mRNA was incubated with ~1 μM PP7ΔFG-mEGFP in a buffer containing 10 mM Tris (pH 7.5), 25 mM NaCl, 0.1 mM EDTA, 0.01 mg/ml yeast tRNA (Ambion), 50 μg/ml heparin and 0.01% (v/v) IGEPAL CA630 for 1 h at room temperature. The sample was then 2x diluted and incubated with 25 nM biotinylated *clpX* capture strand (DHL_P746) and with 1 μM of the five 50-nt long *clpX* FISH probes (200 nM per probe). The incubation was done in buffer containing 2x SSC (Ambion) and 15% (v/v) formamide (Ambion) for 1 hour at 37 ˚C. The sample (50 nM *clpX*-24xPP7 mRNA) was then diluted to a final concentration of 50 pM clpX-24xPP7 mRNA in buffer B (5 mM Tris (pH 8.0), 1 mM EDTA and 1 mM $MgCl_2$) with 1 mg/ml BSA (Sigma-Aldrich). Surface-immobilized mRNAs were imaged in a double-sided sticky tape flow chamber with an epi-fluorescence microscope (see below). No 1.5 coverslips were sequentially sonicated for 30 min in 1 M KOH, acetone, 100% ethanol and ddH_2O and 3x rinsed with ddH_2O between sonications. The flow chamber channels (always 20 μl volume per channel) were incubated for 2 min with 1 mg/ml BSA-Biotin (Sigma-Aldrich), washed with buffer A (10 mM Tris (pH 8.0) and 100 mM NaCl), incubated for 2 min with 0.5 mg/ml Streptavidin (Sigma-Aldrich), washed once with buffer A and twice with buffer B, incubated for 2 min with the sample (50 pM *clpX*-24xPP7 mRNA), washed twice with buffer B and sealed with 5-min epoxy.

Surface-immobilization and imaging of single GFP molecules

The purified iAPmEGFP-SNAPf protein was obtained from Burak Okumus (Paulsson Lab, Harvard Medical School). The protein was diluted 1:400 in 1x PBS with 1 mg/ml BSA (Sigma-Aldrich) and surface-attached and imaged as described above for the *in vitro clpX* FISH experiment. The histograms of GFP spot intensities and photo-bleaching step sizes were fitted with a single Gaussian $y = a_1\, e^{-\frac{1}{2}\left(\frac{x-b_1}{c_1}\right)^2}$. The equation has three free fitting parameters (i.e. scaling parameters a_1, mean b_1 and standard deviation c_1).

1000 random positions were analyzed to estimate the GFP background. The mean fluorescence intensity in a 7 x 7 box around the random position was calculated and the resulting histogram was fitted as described above for the histogram of GFP spot intensities. For the photo-bleaching step analysis, 473 single-molecule traces were manually inspected and rejected from the analysis if (i) the traces were too noisy, (iii) showed more than one photo-bleaching step or (iii) severe blinking. 138 traces (29.2%) were rejected. The remaining traces were fitted to two horizontal lines with the fminsearch function in Matlab. The first line corresponds to the fluorescent state and the second line to the bleached state.

Epi-fluorescence microscopy

The epi-fluorescence microscopy experiments were performed on an inverted microscope (Nikon Ti-E) equipped with a Perfect Focus System (PFS, Nikon), an Orca-R2 (Hamamatsu) camera, an automated x-y stage (BioPrecision2 Inverted Stage, Ludl) and a 100x Plan Apo objective with numerical aperture (NA) of 1.4 (Nikon). Image acquisition was controlled by Micro-manager (http://valelab.ucsf.edu/~MM/MMwiki/).

Fluorescence imaging was performed with an LED system (Spectra 7 light engine, Lumencor) and the following filter cubes: for DAPI, LF405-A (Semrock); for GFP, GFP-3035B (Semrock); and for Cy3, TRITC-A (Semrock). Individual GFP molecules were imaged with 2-s exposure time (cyan LED) and no binning. Individual Cy3 and Quasar570 dyes were imaged with 2-s exposure time (green LED) and no binning. The fluorescence images were acquired without binning (64.5 nm effective pixel size) and saved as 16-bit TIFF files.

Highly inclined and laminated optical sheet (HILO) microscopy

E. coli cells were grown at 30 °C in imaging medium (M9 medium with 10% (v/v) LB medium) to exponential phase and cells were imaged on an agar pad made of 2% (w/v) low-melting agarose (QA-Agarose, MP Biomedicals) dissolved in imaging medium. The microscopy was performed at room temperature. Bacteria were imaged via an in-house built objective-type TIRF-based configuration on a modified inverted microscope (Nikon Ti-E) equipped with a 100x Plan Apo objective with a numerical aperture (NA) of 1.4 (Nikon) and a 2.5x C-mount adaptor (Nikon). Illumination was achieved by a 488-nm high-power optically pumped semiconductor (OPSL) laser (Genesis MX488-1000 STM, Coherent) in HILO[131] mode using 2 mW intensity. The incidence angle was adjusted empirically to optimize for the best signal-to-noise ratio. Images were acquired using an electron-multiplying charge-coupled device (EMCCD) camera (ixon3 897, Andor) with EM gain set to 300 and a 5x preamplifier gain. The camera was controlled by Micro-manager (http://valelab.ucsf.edu/~MM/MMwiki/). The effective pixel size of the acquired images corresponded to 64 nm (250x magnification). For green fluorescence imaging, a dichroic filter (Di01-R488, Semrock) and a bandpass emission filter (FF01-525/45, Semrock) were mounted on a custom-made plastic filter cube (designed and

manufactured by Burak Okumus). The exposure time for GFP was 30 ms and images were acquired with a ~15 Hz frame rate.

Single-molecule mRNA FISH (Fluorescence *in situ* hybridization)

The FISH protocol for *E. coli* was adapted from existing protocols for *Saccharomyces cerevisiae*[212,219]. In short, overnight *E. coli* cultures (MC4100 wild type and DHL708 as negative control) were diluted 1:1000 in LB medium (v_{tot} = 50 ml) and the cells were grown at 37 ˚C until an OD_{600} of 0.5. The cells were fixed by adding 3.7% (v/v) formaldehyde (Ted Pella, 16% stock) to the culture and incubated for 45 min at room temperature in a shaker (180 rpm). The cells were then pelleted (3200 g, 10 min, 4 ˚C) and washed three times with ice-cold SP buffer (30 mM Tris-HCl (pH 8.0) 23% (v/v) sucrose). The cell pellet was resuspended in 1 ml SP buffer and transferred into three Eppendorf tubes (~300 µl per tube) to test different lysozyme conditions (the different test conditions are not described here). One of the Eppendorf tube with 300 µl cells was then pelleted (4000 g, 2 min, 4 ˚C) and washed with 1 ml SP buffer. The pellet was resuspended in 1 ml SP buffer with 50 µg/ml lysozyme and 5 mM EDTA and the reaction was incubated for 30 min at room temperature. The cells were pellet (4000 g, 2 min, 4 ˚C), washed twice with 1 ml SP buffer, and 200 µl cell suspension was pipetted onto a 22 x 22 mm poly-L-lysine (PPL) coated coverslip. The coverslip was cleaned as described (see above) and coated with PLL by pipetting 200 µL of 0.01% (w/v) PLL solution on the coverslip and incubated for 10 min at room temperature. The PLL solution was then removed with a suction and the coverslip was covered with a lid and air-dried for > 30 min. The coverslip was rinsed 3x with 1 ml ddH_20 and again air-dried for > 30 min. The cells were incubated on the coverslip for 60 min at room temperature. The coverslip was then transferred into a 6-well dish and washed with 2 ml SB buffer to wash away unbound cells. The SB buffer was removed with a vacuum suction and 2 ml

70% (v/v) EtOH (stored at –20 ˚C) was added to the well to permeablize the cell membrane. The plate was incubated in a –20 ˚C freezer for o/n. After the overnight incubation, the plate was briefly warmed up to room temperature and the EtOH was exchanged against 2 ml 2x SSC (Ambion) to rehydrate the cells for 5 min at room temperature. The solution was exchanged against 2 ml 40% (v/v) formamide (Ambion) 2x SSC and the hybridization mix was prepared. All work with the FISH probes was done under low light conditions to prevent bleaching of the fluorescent dyes. Solution F (40% (v/v) formamide, 2x SSC, 1x PBS, 0.5 µg/µl ssDNA, 0.25 µg/µl N50 probe and 0.1 ng/µl FISH probes) was incubated at 95 ˚C for 3 min and cooled down to room temperature after the heating. The N50 probe is a mixture of many 50-bp-long oligos with random sequences. 20 µl of solution F was then mixed with 20 µl of solution H (40% (v/v) formamide, 2x SSC, 2 mg/ml BSA and 10 mM ribonucleoside-vanadyl complex (NEB)), pipetted on a piece of parafilm and the coverslip (with the surface that has the cells attached facing down) put on the drop of hybridization solution. The sample was then sealed (to prevent evaporation) and incubated for ~14 h at 37 ˚C in the dark. The coverslip was put back into the 6-well plate after the overnight incubation and washed 2x with 2 ml 40% (v/v) formamide 2x SSC with 15 min incubations at 37 ˚C between washes. The coverslip was incubated in 2 ml 2x SSC 0.1% (v/v) Triton-X for 15 min at room temperature followed by a 15-min incubation in 2 ml 2x SSC at room temperature. The coverslip was then incubated in 2 ml 1x PBS with DAPI for 2 min and washed twice with 2 ml 1x PBS at room temperature. The coverslip was dipped into 100% (v/v) EtOH, briefly air-dried (~5 min) and mounted with ProLong Gold (Invitrogen) antifade reagent. Cells were imaged with a Nikon Ti-E epi-fluorescence microscope as described above. Z-stacks (16 planes, 0.2 µm spacing) were taken for Cy3 and single images at the cell center for DAPI. Maximum-projections of the Cy3 z-stacks were generated used for the spot analysis. Spots were identified and quantified as described for the *in vitro* FISH

analysis (see above). The histogram of Cy3 spot intensities of the Δ*clpX* strain was fitted with the sum of two Gaussian using following formula: $y = a_1\, e^{-\frac{1}{2}\left(\frac{x-b_1}{c_1}\right)^2} + a_2\, e^{-\frac{1}{2}\left(\frac{x-b_2}{c_2}\right)^2}$ where $b_2 = 2b_1$ and $c_2 = 2c_1$. The equation has 4 free fitting parameters (i.e. two scaling parameters a_1 and a_2, mean b_1 and standard deviation c_1). The histogram of Cy3 spot intensities of the wild-type strain was fitted with the sum of three Gaussian using $y = a_1\, e^{-\frac{1}{2}\left(\frac{x-b_1}{c_1}\right)^2} + a_2\, e^{-\frac{1}{2}\left(\frac{x-b_2}{c_2}\right)^2} + a_3\, e^{-\frac{1}{2}\left(\frac{x-b_3}{c_3}\right)^2}$ where $b_2 = 2b_1, b_3 = 3b_1, c_2 = 2c_1$ and $c_3 = 3c_1$. This equation has 5 free fitting parameters (i.e. three scaling parameters $a_1 - a_3$, mean b_1 and standard deviation c_1).

3.5. Appendix

Investigation of the Spinach aptamer for imaging low-to-medium abundant mRNAs in *E. coli* cells

I tagged the *clpX* mRNA with the Spinach aptamer to test Spinach for live-cell imaging purposes. DFHBI is a small molecule that is non-fluorescent and becomes fluorescent upon binding to the Spinach aptamer[210]. We confirmed this effect by performing bulk measurements of the Spinach RNA aptamer with DFHBI *in vitro* (data not shown, experiment performed by collaborators Ralf Jungmann and Maier Avendano). But we also observed that the DFHBI dye produces a very strong signal when added to wild-type *E. coli* cells that lack the Spinach aptamer (**Figure 3.10**). This is probably due to non-specific activation of the DFHBI dye by endogenous *E. coli* RNAs. The fluorescent signal increases by a very small amount when the DFHBI dye is added to cells that express the *clpX* mRNA tagged with the Spinach aptamer.

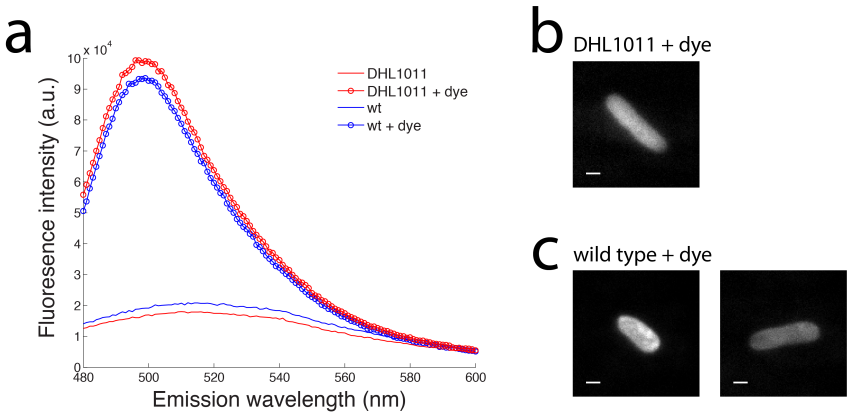


Figure 3.10. The DFHBI dye of the Spinach aptamer cannot be used for the detection of low-to-medium abundant mRNAs in *E. coli*. (**a**) Addition of the DFHBI dye to a cell suspension of wild-type (wt) cells and to *E. coli* cells, which have the Spinach aptamer fused to the 3' end of the *clpX* mRNA (strain DHL1011), results in a strong signal (measured with a fluorometer in bulk). The signal appears marginally stronger for the strain with the Spinach aptamer (compare 'DHL1011 + dye' to 'wt + dye') but the non-specific signal is very high (compare 'wt' to 'wt + dye'). The same qualitative result was obtained when single cells with (**b**) or without the Spinach aptamer (**c**) were imaged in the presence of the DFHBI dye. Two different images are shown for the wild-type control. The scale bars are 1 µm.

Thus, we concluded that Spinach cannot be used for the detection of low-to-medium abundant mRNAs in *E. coli* and is very far away from single-molecule sensitivity. The Spinach aptamer could potentially still be of useful for measuring the dynamics of mRNAs in astronomically high numbers but this was not our aim. We therefore decided not to use Spinach.

3.6. Description of Movies

Video 3.1

Live-cell HILO microscopy of cells co-expressing the tagged *clpX*-24xPP7 transcripts and the PP7-mEGFP fusion protein (strain DHL1177). The movie shows discrete PP7-mEGFP particles moving around in the cell. The particles like correspond do individual tagged *clpX* mRNAs but further experiments are necessary to prove this unequivocally. Scale bars (white) are 1 µm. The images of the movie sequences were subjected to a quantitative grayscale scaling (left) and to a 'per-frame auto-scaling' (right) to better display the particle movement despite fast photobleaching.

Video 3.2

Live-cell HILO microscopy of wild-type cells expressing the PP7-mEGFP fusion protein (strain DHL1203). The green fluorescence signal displays a uniform cytoplasmic distribution. Scale bars (white) are 1 µm. The images of the movie sequences were subjected to a quantitative grayscale scaling (left) and to a 'per-frame auto-scaling' (right) to better display the particle movement despite fast photobleaching.

4. The RpoS-SprE negative feedback loop

4.1. Introduction

4.1.1. Negative feedback loops and homeostasis

Negative feedback plays many important roles in biological systems, from providing homeostasis in the face of external perturbations[220] to inducing stable oscillations[96]. It has also been suggested that negative feedback suppresses the spontaneous noise that arises from probabilistic chemical reactions between molecules present in low numbers. For example, engineering transcriptional negative feedback loops in *Escherichia coli* (*E. coli*) can reduce such noise[107,221,222], while eliminating the natural negative feedback loops controlling replication of bacterial plasmids can increase the noise levels dramatically[223], which typically leads to higher plasmid extinction rates.

Negative feedback loops adjust the rates of synthesis or degradation to correct changes in molecular abundances, but the adjustments are only guaranteed to significantly suppress fluctuations if the deviations strongly and directly affect the rates of synthesis or degradation[224]. These conditions are rarely satisfied since lags and delays are omnipresent, chemical reactions are inherently probabilistic, and many systems are restricted to feedback loops with low effective cooperativity. Thus negative feedback may either suppress or amplify fluctuations depending on the exact molecular mechanism and parameters. However, for some types of mechanisms, effective noise suppression can be effectively ruled out. For example, a recent theoretical study[17] showed that there are severe limits on how well any negative feedback control loop could suppress molecular fluctuations when mutual information is lost due to stochastic signaling. Unpublished work (J. Paulsson, personal communications) shows that

similarly severe limits can be demonstrated for systems with lags and delays, or systems with limited cooperativity.

In many practical applications it is also hard to infer the selective value of fluctuations, for example if some protein fluctuations are harmful, neutral or beneficial for the organism. This makes it hard to evaluate the feedback loop design. For this reason I chose to study the naturally occurring RpoS-SprE negative feedback loop controlling bacterial stress response and stationary phase, where variability is known to have great physiological consequences. RpoS controls hundreds of genes in *E. coli* and is a key player in everything from pathogenicity[225] to nutrient starvation[8,9]. However, despite the fact that RpoS is one of the most studied proteins in *E. coli*, its self-controlling feedback loops have not yet been quantified due to various technical challenges.

We started to study this system in collaboration with C. Peterson (Suffolk University) and T. Silhavy (Princeton University), who originally discovered the feedback loop[11] and measured some of the parameters. Using these parameters (and others from the literature) in mathematical models suggested that this negative feedback loop in fact cannot suppress spontaneous fluctuations due to low numbers of molecules: the lags are too long, the cooperativity is too low, and the signaling molecules are made at extremely low rates. In such cases, even *negative* feedback loops are expected to amplify rather than suppress random fluctuations, unless fluctuations come from slowly changing external inputs, such as changes in conditions. This observation motivated the mathematical and experimental research presented here, which is focused on determining what role this feedback loop plays in stress response.

4.1.2. Bacterial stress response

RpoS (also known as σ^S) is an alternative sigma factor present in a group of gram-negative bacteria including the model organism *E. coli*. It serves as a master regulator of the general stress response[8] and replaces the vegetative sigma factor σ^{70} from the RNA polymerase under a wide range of stressful conditions. RpoS in turn redirects the transcriptional program of ~500 genes[7] towards stress resistance and damage repair[8,226], which retards cellular senescence[227] and renders the bacterium broadly stress-resistant[228].

RpoS integrates and responds to multiple signals from the environment, including a lack of nutrients, low pH, high osmolarity, UV, oxidative stress and toxic chemicals, which control RpoS at the levels of transcription, translation, activity and protein degradation[8,9] (**Figure 4.1**).

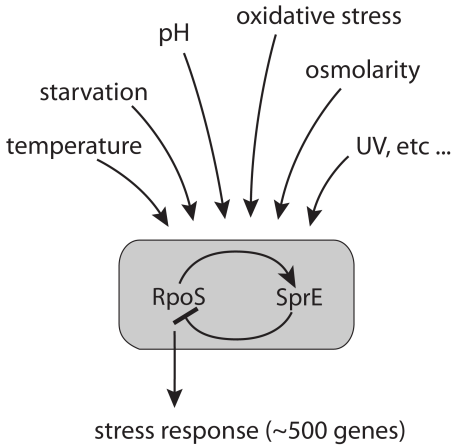


Figure 4.1. The *E. coli* stress response sigma factor RpoS responds to signals from the environment and is subjected to internal control by a complicated regulatory network, which is represented here by the RpoS-SprE negative feedback loop.

During cell growth, RpoS is also subject to self-control via an indirect negative feedback loop that regulates RpoS degradation[10,11,16]: RpoS promotes the transcription of the *sprE*

(also known as *rssB*) gene[10,16], while the adapter protein SprE in turn targets RpoS for degradation by the ClpXP protease[229,230] (**Figure 4.2**).

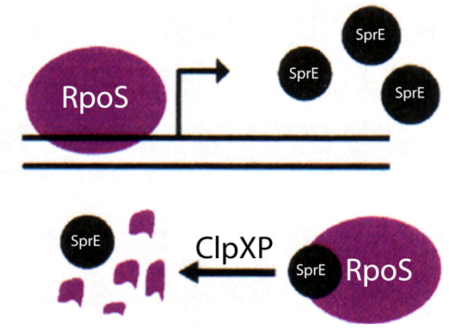


Figure 4.2. A negative feedback loop by regulated degradation is believed to be the core process of RpoS regulation in exponential-phase *E. coli* cells. The sigma factor RpoS activates transcription of the *sprE* gene (core polymerase not drawn). The adapter protein SprE targets RpoS for degradation, thereby closing the negative feedback loop.

Thus, the negative feedback loop is composed of one transcription arm (RpoS activates transcription of *sprE*) and one protein-protein interaction arm (SprE delivers RpoS to the ClpXP protease). This is a very common network motif occurring in diverse systems such as p53-Mdm2[3,12], NFκB-IκB[13], Ime1-Ime2[231], Smo-Ptc[232] and the heat-shock response[14,15].

4.1.3. Regulation of RpoS in exponential- and stationary-phase cells

RpoS levels are very low in exponential phase because of fast and efficient SprE-mediated proteolysis[233-236]. Recognition of RpoS by SprE and ClpX is sequential: SprE binds to an internal region of RpoS and forms a 1:1 complex. Upon SprE binding, the N-terminal, cryptic ClpX-binding site of RpoS gets exposed, leading to a ternary complex (ClpX-RpoS-SprE)[237,238] that is recognized by the protease ClpP[229,239]. Both binding sites are necessary for RpoS degradation; ClpX does not interact with SprE alone[237]. The

SprE level determines the rate for RpoS degradation in exponential phase. For example, elevating the cellular level of SprE increases RpoS degradation[10]. RpoS binding to SprE or RNA polymerase is mutually exclusive; the sigma factor is protected from degradation when bound to the polymerase[240]. Normally, sigma factors only engage briefly with the RNA polymerase to initiate transcription and then dissociate during transcript elongation.

As cells enter stationary phase, RpoS degradation ceases despite high SprE levels, causing RpoS to accumulate[8,10,11,241] to ~170–230 molecules per cell[234], with an increase in half-life from ~1.5 min to over 30 min[229,230,233,235]. This is partly because carbon starvation increases translational errors during protein synthesis, resulting in aberrant proteins that sequester the ClpP protease, competing with RpoS for degradation[227]. RpoS degradation is also very sensitive to ATP levels; it was recently shown that ClpXP cannot degrade RpoS *in vitro* at low ATP availability[242]. This suggests that RpoS levels are also directly controlled by intracellular ATP levels and hence nutrient availability[242].
The RpoS-SprE system is often envisioned as a genetic switch in the OFF state in exponential phase (low RpoS) and in the ON state (high RpoS) under stressful conditions. An acute stress could then flip the switch from OFF to ON (**Figure 4.3**).

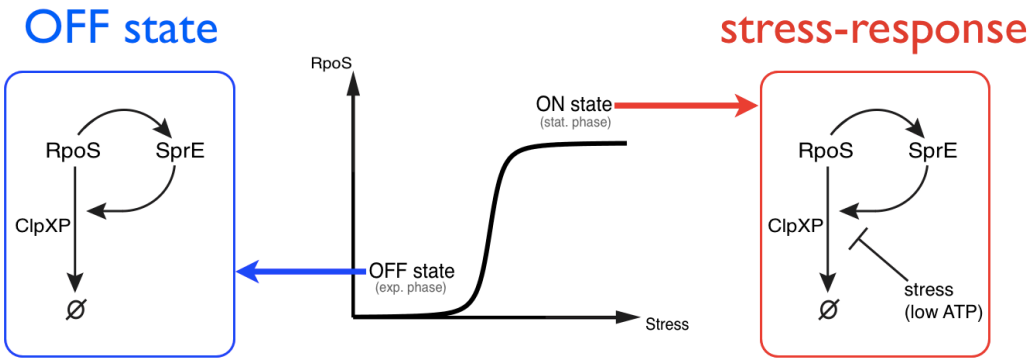


Figure 4.3. RpoS is present in low numbers and is short-lived in exponential-phase cells (OFF state). The RpoS level is believed to be controlled by a negative feedback loop by regulated degradation. RpoS activates transcription of the *sprE* gene and SprE in turn delivers RpoS to the ClpXP protease for degradation. When cells are exposed to a stress, RpoS stabilizes because some other component interferes with SprE or the ClpXP protease. This causes RpoS to accumulate to high levels (ON state) and redirect the transcription machinery towards stress response genes.

4.1.4. Suggested roles for the negative feedback control of RpoS

It has been suggested that the function of the RpoS-SprE negative feedback loop is to maintain homeostasis and decrease noise in RpoS. The argument is that any spontaneous fluctuation in the environment or random RpoS synthesis events that cause an increase in RpoS should lead to a simultaneous increase in SprE, which in turn should reduce RpoS to its initial level by adjusting the RpoS degradation rate. Hence the control loop would maintain RpoS level within a narrow appropriate range[8,10,11,16]. Pruteanu and Hengge-Aronis speculated that the role of the homeostasis is to stabilize the RpoS OFF state and keep cells from entering stationary phase by mistake[10]. Besides the homeostatic feedback control, it was suggested that high levels of SprE in stationary-phase cells are important so that the cells can exit stationary phase quickly after conditions changed[10,11,16].

4.1.5. Our approach

The central aim of this project is to quantitatively study the RpoS-SprE feedback loop in individual *E. coli* cells and attempt to explain how it could facilitate stress response. We also hope that analyzing this particular system will provide more general insights into how cells may use feedback control loops.

First, I performed mathematical analysis of the RpoS-SprE negative feedback loop using deterministic and stochastic modeling (see Section 4.2). In brief, the stochastic approach suggests that the negative feedback loop would increase single-cell variability in the cellular RpoS level if the noise came from the low levels of RpoS or SprE, but would suppress variability if the noise came from slowly varying changes in the parameters of the system. The deterministic analysis in turn suggests that the unusual parameters and design of the RpoS-SprE negative feedback should allow a faster target gene response when the cells are exposed to a stress.

I then set out to evaluate these predictions experimentally by analyzing the RpoS feedback loop dynamics in single *E. coli* cells. The ultimate goal is to observe the feedback loop dynamics directly by tagging the RpoS and SprE proteins with two spectrally distinct fluorescent proteins and simultaneously quantify the RpoS and SprE protein levels in single cells. However, one reason that so little is known about this system is that the key proteins do not readily accept tags. There have already been many attempts to tag RpoS in various ways, but no such tags retained any significant activity of the molecule. RpoS is also degraded on a time scale of ~1 minute in exponentially growing cells, such that fluorescent protein reporters would not have sufficient time to mature. Building and evaluating useful reporters was therefore a significant aim in its own right.

4.2. Theory

4.2.1. Negative feedback loops can amplify noise

Rather than attempting large-scale computer simulations where many parameters and interactions are guessed, I constructed some simple toy models that primarily are intended to promote intuition. For simple toy models, fluctuations in biological systems can be analyzed with a fluctuation-dissipation type of relation for linearized non-equilibrium systems[218,243]. The simplest model of the negative feedback loop by regulated degradation can be described by the following birth and death processes:

$$\begin{matrix} x_1 \xrightarrow{\lambda_1} x_1 + 1 \\ x_1 \xrightarrow{\lambda_2 x_1 x_2} x_1 - 1 \end{matrix} \quad \text{and} \quad \begin{matrix} x_2 \xrightarrow{\lambda_3 x_1} x_2 + b \\ x_2 \xrightarrow{\lambda_4 x_2} x_2 - 1 \end{matrix}, \tag{1}$$

where x_1 is the number of RpoS molecules, x_2 is the number of SprE molecules and λ_i are rate constants for the transition probabilities for the various events. This description condenses the mRNA dynamics for SprE into a transcriptional burst (of size b), a non-perfect approximation that greatly simplifies the algebra involved. For RpoS, fluctuating production rates are instead analyzed in greater detail below. Because both RpoS and SprE are present in extremely low numbers – often reported as too low to measure – it is plausible that the fluctuations actually do come from having low numbers of the proteins themselves.

We then assume that the random process can be approximated as stationary, in the sense that levels can fluctuate up and down in individual cells, but the overall distribution across the population has stabilized. This approximation may seem odd since the system governs the transition from exponential to stationary phase, an inherently non-stationary process. However, the half-lives of the components are on the order of a few

minutes, and the process is therefore expected to equilibrate to quasi-stationary states as the parameters change during growth.

The result obtained from this approach can generally be expressed in terms of average lifetimes (i.e. τ_i) and average numbers of molecules (i.e. $\langle x_i \rangle$) at the stationary state, which implicitly accounts for all parameter values. The normalized variance in RpoS levels then follows as:

$$\eta_{11_{closed-loop}} = \frac{\sigma_1^2}{\langle x_1 \rangle^2} \approx \overbrace{\frac{1}{\langle x_1 \rangle}\left(1 - \frac{1}{2}\frac{\tau_1}{\tau_1 + \tau_2}\right)}^{Intrinsic\ noise} + \overbrace{\frac{1}{2}\frac{\langle b \rangle}{\langle x_2 \rangle}\frac{\tau_2}{\tau_1 + \tau_2}}^{Extrinsic\ noise}, \quad (2)$$

where τ_1 and τ_2 are the average lifetimes at steady state of x_1 and x_2, respectively. An open-loop system without negative feedback by regulated degradation (i.e. degradation of x_1 is independent of x_2) instead exhibits Poisson fluctuations with $\eta_{11_{\mathrm{open-loop}}} = \sigma_1^2/\langle x_1 \rangle^2 = 1/\langle x_1 \rangle$.

Equation 2 shows that negative feedback can suppress the noise in component x_1 (η_{11}) effectively if the inhibitor x_2 is present in large numbers and has a much shorter lifetime (τ_2) than that of the controlled molecule x_1 (τ_1), i.e. $\langle x_1 \rangle \ll \langle x_2 \rangle$ and $\tau_1 \gg \tau_2$. This conclusion was confirmed with systematic Gillespie simulations for different parameter values (**Figure 4.4**). If the RpoS-SprE negative feedback loop evolved to reduce such low-copy noise, then its parameters should thus be expected to follow this condition. The relative protein levels and half-lives of the sigma factor RpoS (x_1) and its 'inhibitor' SprE (x_2) were measured in exponential phase. Surprisingly, it was found that the ratio between RpoS and SprE levels is approximately 10:1 (i.e. $\langle x_1 \rangle \gg \langle x_2 \rangle$); the authors estimated that only a few molecules of SprE are present per cell[10,241]. The SprE protein level is in turn below Western blot detection limit in exponential phase. Overexpression

of SprE allowed half-life measurements and showed that the SprE protein half-life was independent of growth phase, and was approximately 30 min[16,244]. The half-life of RpoS was determined to be ~1.5 min[229,230,233,235] – one of the most unstable proteins measured in *E. coli*. Thus the parameters are in the opposite extreme compared to what is expected for efficient RpoS noise suppression: rather than having short-lived abundant signals, the feedback loop relies on a long-lived low-abundance signal. In fact, the theory for hard limits on feedback control[17] show that, regardless of the specific assumptions and approximations used here, such systems cannot suppress spontaneous fluctuations.

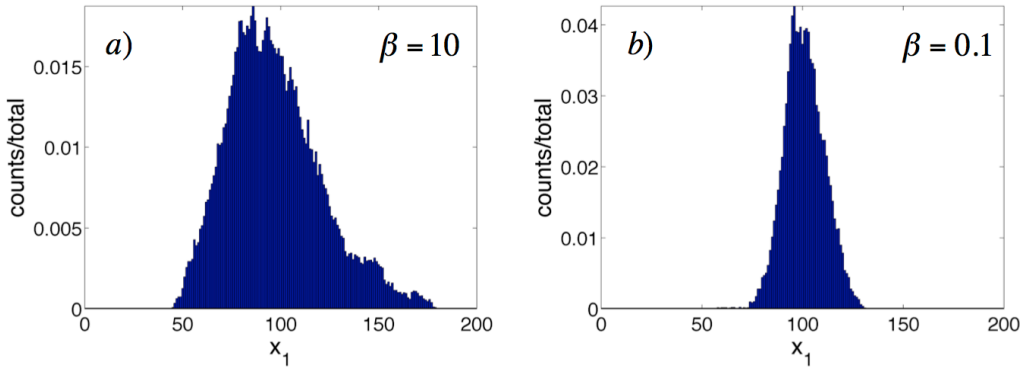


Figure 4.4. The normalized histogram of x_1 for (**a**) a system with a long-lived inhibitor molecule, x_2 ($\tau_1 = 0.1$ and $\tau_2 = 1$, as in the wild-type RpoS-SprE system) and (**b**) a system with a short-lived inhibitor molecule, x_2 ($\tau_1 = 1$ and $\tau_2 = 0.1$). The parameter β is $\beta = \tau_2/\tau_1$. The average copy numbers for $x_1$and x_2 are in both systems $x_1 = 100$ and $x_2 = 10$, respectively. The burst size is $b = 1$.The Gillespie simulations were run for 100,000 steps. System a) has a larger noise of x_1 than system b): $\eta_{11_{\text{system a}}} = 0.062 > \eta_{11_{\text{system b}}} = 0.0095$.

To account for upstream sources of noise in RpoS expression rates I use an extended three-variable model where RpoS synthesis λ_1 depends on a fluctuating environmental variable, x_0. For any model in which x_0 autocorrelations decay exponentially, regardless of bursting or other details of the x_0 process, we can again calculate the normalized

variance. The full solution is a complicated expression, but based on the measured parameters ($\tau_1 \ll \tau_2$) and assuming that the environment has a slow correlation time compared to the geometric average of RpoS and SprE ($\sqrt{\tau_1 \tau_2} \ll \tau_0$), which is likely because both proteins are so quickly degraded, it follows that:

$$\eta_{11_{RpoS-SprE}} = \frac{\sigma_1^2}{\langle x_1 \rangle^2} \approx \overbrace{\frac{1}{\langle x_1 \rangle}}^{\substack{Intrinsic\ noise \\ (low\ RpoS)}} + \overbrace{\overbrace{\frac{1}{2}\frac{\langle b \rangle}{\langle x_2 \rangle}}^{low\ SprE} + \overbrace{\eta_{00}\frac{\tau_0 + 2\tau_2}{4\tau_0 + 2\tau_2}}^{\substack{Environmental\ noise \\ (time-averaged)}}}^{Extrinsic\ noise} \quad (3)$$

Analyzing this expression shows that, within the model, the system fails to decrease spontaneous noise in RpoS, introduces large fluctuations from SprE, but to some extent suppresses slow environmental noise. However, even for the slowest environments, the relative standard deviation could only be reduced twofold compared to the equivalent open-loop system. The poor noise suppression capability directly reflects the long relative lag and the statistical uncertainty in the feedback response.

A similar feedback circuit was modeled in great detail for the heat-shock sigma factor (σ^H) in *E. coli*. The authors by contrast arrived at the conclusion that the σ^H negative feedback loop generates faster and smaller fluctuations[15], as commonly believed for negative feedback. The difference lies in the specific parameters for the two systems under study. In case of the heat shock negative feedback loop, the 'inhibitor' was assumed to be short-lived and present in high-numbers, in direct contrast to the negative-feedback regulator SprE.

The prediction that the RpoS-SprE system functions to increase fluctuations may seem non-biological at first, but large fluctuations in a stress response system could very well be advantageous for bacterial populations in the wild. The problem is that plausible

arguments could be made both for bet hedging – suggesting that the population as a whole could benefit from a few cells being prepared for stress – and for homeostasis, to insure that cells do not accidentally go into stationary phase and forsake an opportunity to divide further. The theory can therefore only be used to clarify basic principles, not make plausible predictions about the actual dynamics, which instead must be measured.

4.2.2 RpoS can overshoot its steady state leading to a faster target gene response

We also analyze the dynamic response to perturbations. This could be done by simulating the same model as above, but to reduce the number of free parameters and allow for analytical solutions without linearizations we instead formulate a deterministic equivalent.

The deterministic behavior of the RpoS-SprE negative feedback loop and a target gene product in exponential phase can be captured by three coupled ordinary differential equations (ODEs):

$$\frac{dx_1}{dt} = \lambda_1 - \beta_1 x_1 x_2, \frac{dx_2}{dt} = \lambda_2 x_1 - \beta_2 x_2 \text{ and } \frac{dx_3}{dt} = \lambda_3 x_1 - \beta_3 x_3, \quad (4)$$

where x_1 is RpoS, x_2 is SprE, and x_3 is now the product of an RpoS-regulated gene. Parameters λ_i and β_i are synthesis and degradation rate constants respectively. The equations of the open-loop mutant system are identical except for the first equation:

$$\frac{dx_1}{dt} = \lambda_1 - \beta_b x_1 \text{ with } \beta_b = \beta_1 \cdot x_{2_{ss}} = \sqrt{\frac{\beta_1 \cdot \lambda_1 \cdot \lambda_2}{\beta_2}}. \quad (5)$$

In this parameterization, x_1, x_2 and x_3 have the same steady state $\left(x_{1_{ss}} = \frac{\lambda_1}{\beta_b}, x_{2_{ss}} = \frac{\lambda_1\lambda_2}{\beta_b\beta_2} \text{ and } x_{3_{ss}} = \frac{\lambda_1\lambda_3}{\beta_b\beta_3}\right)$ and average lifetime at steady state $\left(\tau_1 = \frac{1}{\beta_b}, \tau_2 = \frac{1}{\beta_2} \text{ and } \tau_3 = \frac{1}{\beta_3}\right)$. The rescaled equations for the RpoS-SprE wild-type system are: $\frac{dx_1'}{d\tau} = 1 - \beta^2 x_2' x_1'$, $\frac{dx_2'}{d\tau} = x_1' - x_2'$ and $\frac{dx_3'}{d\tau} = x_1' - \alpha x_3'$ where the rescaled variables are indicated with a prime $\left(\frac{dt}{d\tau} = \frac{1}{\beta_2}, x_1' = \frac{\beta_2}{\lambda_1} x_1, x_2' = \frac{\beta_2\beta_2}{\lambda_1\lambda_2} x_2 \text{ and } x_3' = \frac{\beta_2\beta_2}{\lambda_1\lambda_3} x_3\right)$. The rescaled equations of the open-loop mutant system are identical except that the first equation is $\frac{dx_1'}{d\tau} = 1 - \beta x_1'$. Solving for the steady state of the rescaled variables results in $x_{1_{ss}}' = x_{2_{ss}}' = \frac{1}{\beta} \text{and } x_{3_{ss}}' = \frac{1}{\alpha\,\beta}$. Hence, rescaling reduced the complexity to two parameters: $\alpha = \frac{\tau_2}{\tau_3}$ and $\beta = \frac{\tau_2}{\tau_1}$.

Simulation of this deterministic three-variable system shows that the negative feedback loop by regulated degradation allows RpoS to activate a target gene response faster than without the feedback control (**Figure 4.5**), particularly when SprE is more long-lived than RpoS. This is because RpoS overshoots its steady state in the wild-type system for large β and hence causes a faster increase of the target gene product, whereas RpoS does not overshoot its steady state in the open-loop mutant. In fact, the system has three features that promote rapid response: a short life-time of RpoS, which sets the general time scale of the response, a negative feedback loop that will speed up adjustments, and a lag in the loop that will allow an over-shoot.

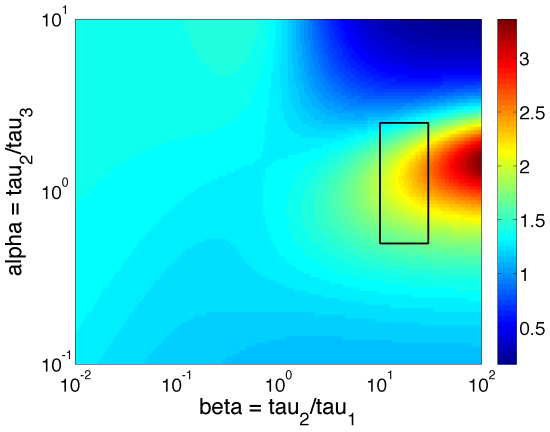


Figure 4.5. Comparison of the target gene response in the open-loop mutant to the wild-type RpoS system. Each system was simulated until the target gene product ($\boldsymbol{x_3'}$) adjusted to its steady state from the initial condition of $\boldsymbol{x'_{1_0} = x'_{2_0} = 0.1\, x'_{1_{ss}}}$ **and** $\boldsymbol{x'_{3_0} = 0}$. This adjustment was quantified by using the 'area under the curve' measure (Material and Methods). The color corresponds to the ratio of the $\boldsymbol{x_3'}$ adjustment of the open-loop mutant and the wild-type RpoS system. Red indicates that the target gene product $\boldsymbol{x_3'}$ of the RpoS wild-type system adjusts faster to its steady state than the open-loop mutant; blue indicates that the target gene product $\boldsymbol{x_3'}$ of the open-loop system adjusts faster. The box (black) represents $\boldsymbol{\alpha}$ values between 0.5 and 2.5, and $\boldsymbol{\beta}$ values between 10 and 30.

The open-loop mutant system adjusts faster only for very large α and β values. This is because a large α value corresponds to a target gene product with a shorter average lifetime at steady state compared to SprE (i.e. $\tau_2 > \tau_3$) causing the target gene product to follow the initial overshoot in RpoS, thereby delaying its steady state adjustment. The wild-type RpoS system shows better adjustment for β values above ~10 and α values between 0.5 and 2.5. Interestingly, this parameter regime corresponds to RpoS-regulated gene products with average lifetimes between 12 and 60 min. This suggests that one reason for this unexpected high β value of the RpoS-SprE negative feedback loop could be the potential ability to activate a target gene response faster than without the feedback control. This could be a useful feature for the cell because it allows a faster response after a sudden stress exposure.

4.3. Experiments

Experimentally evaluating the dynamics of the RpoS-SprE feedback loop is challenging for many reasons. First no group has succeeded in making a functional protein fusion between RpoS and a fluorescent reporter. Both RpoS and SprE are also present in very low numbers during exponential phase, which makes detection with standard epi-fluorescence microscopy challenging. Additionally, the RpoS half-life is very short in exponential phase (~1.5 min) and even the fastest maturing GFP variant, Venus YFP, has an average maturation time of ~7 min[117,245], so most of the Venus-tagged RpoS fusion proteins would be degraded before being detectable.

Even an imperfect fusion could still be very useful for relative comparisons between conditions. Traditionally, the regulation of RpoS degradation has been studied in bulk with a fusion between LacZ and a truncated version of RpoS (i.e. RpoS742-LacZ or RpoS750-LacZ)[229,233,237,239,246-248]. These LacZ reporters include both degradation signals and are degraded with the same properties as wild-type RpoS. Bacterial LacZ fusion proteins are often degraded only partially, resulting in a cleaved fusion protein and stable LacZ, but this is not the case for the RpoS degradation reporter[246].

However, all RpoS-LacZ degradation reporters, even the full length versions, have been non-functional as sigma factors, probably due to the immense size of LacZ which forms a tetramer[249] with a molecular weight of 465 kDa and is even larger than the *E. coli* RNA polymerase (379 kDa). Such a reporter can be used for qualitative analysis, but because of the feedback loop, its statistical properties may differ significantly from the wild type since a spontaneous low-copy fluctuation to higher RpoS-LacZ levels would not cause a feedback response, as it would for the wild type.

4.3.1. The RpoS750-Venus degradation reporter is enriched in a small fraction of exponential-phase cells

Despite its imperfections, I constructed a Venus degradation reporter (i.e. RpoS750-Venus) and integrated this construct (i.e. P_{rpoS_long}-*rpoS*750-Venus) in the *phoA* site of the *E. coli* chromosome (leaving the endogenous *rpoS* gene intact). I observed a small fraction of cells (i.e. ~0.1%) that had high levels of RpoS in late-exponential phase. These cells might populate the RpoS ON state likely because RpoS stabilized due to reduced degradation (**Figure 4.6**).

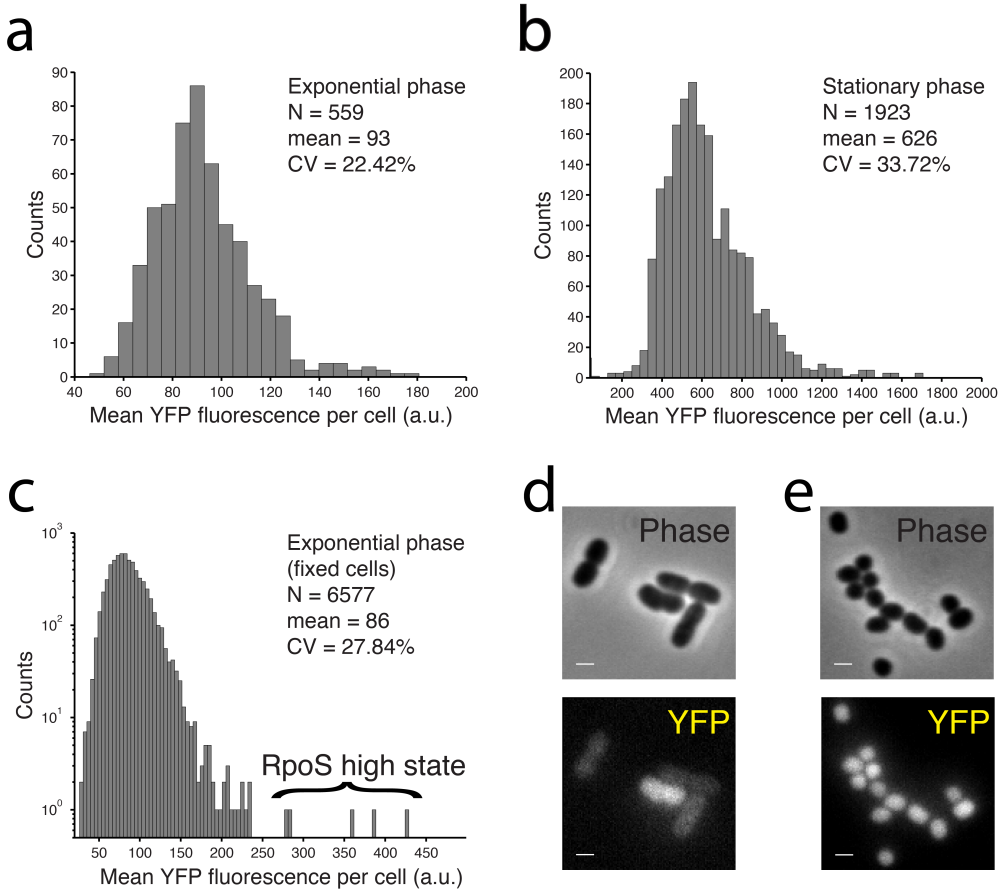


Figure 4.6. The RpoS750-Venus degradation reporter is enriched in a small fraction of cells in exponential phase. This population might correspond to the RpoS ON state. (**a**) Quantification of the cellular RpoS-Venus levels in live cells in exponential phase (OD_{600} = 0.16), (**b**) in live cells in stationary phase (overnight culture) and (**c**) in late-exponential-phase (OD_{600} = 1.0) cells, which were fixed with formaldehyde. (**d**) Yellow fluorescence image of fixed cells in late-exponential phase shows a cell with high levels of RpoS. (**e**) Yellow fluorescence image of live stationary-phase cells shows that cells are small and round in shape, and have high levels of the RpoS750-Venus degradation reporter because RpoS degradation stopped. Scale bars (white) correspond to 1 μm.

This data is consistent with the hypothesis that the RpoS level fluctuates heavily in single cells resulting in a small fraction of cells populating the RpoS ON state, even in non-stress conditions. It will need further experimental validation to show that this small fraction of cells in the RpoS high state is caused by the slow feedback loop dynamics and not by an artifact of the reporter. I did show that the RpoS750-Venus degradation

reporter is actively degraded in mid-exponential-phase cells and that the adapter protein SprE is necessary for the degradation (**Figure 4.7**).

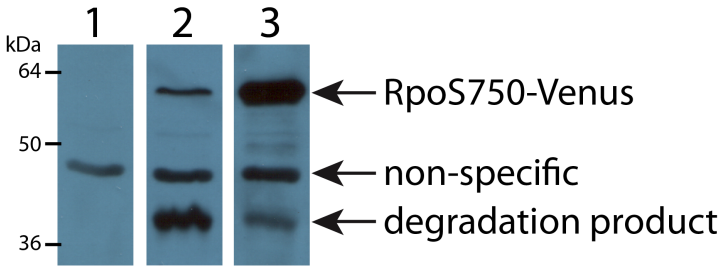


Figure 4.7. Western blot analysis with the αGFP antibody shows that the degradation of the RpoS750-Venus degradation reporter depends on SprE in mid-exponential-phase cells (OD_{600} = 0.7–0.8). The RpoS750-Venus degradation reporter is low in the wild-type background likely because of efficient degradation by the ClpXP protease (2: DHL222). Deletion of SprE (i.e. Δ*sprE* allele) leads to strong accumulation of the RpoS750-Venus reporter (3: DHL363), demonstrating that RpoS750-Venus is actively degraded and that SprE is necessary for degradation. The αGFP antibody cross-reacts with a band in the wild type (1: MC4100). The expected molecular weight of the RpoS750-Venus band is 55.4 kDa.

4.3.2. Evaluation of protein fusions to RpoS

A functional fluorescent protein fusion to RpoS would be the ideal reporter for studying RpoS dynamics in single cells since such a reporter contains more information about the system than the RpoS750-Venus degradation reporter, which is non-functional and can essentially only be used in an RpoS wild-type background. The RpoS750 reporter might still capture some of the RpoS dynamics but a non-functional RpoS reporter should for example have no correlations with SprE in single cells, and due to the feedback, this means that not even the RpoS fluctuations should be accurate.

I therefore constructed a diverse collection of GFP fusions to RpoS (see Materials and Methods) and evaluated them by measuring RpoS activity using a *sprE* promoter (i.e. an RpoS-regulated promoter) fusion, which activity highly depends on RpoS, and

measuring the promoter activity in both the wild type and the RpoS fusion strain. If the fusions affect the levels or activity of RpoS, the downstream promoter activity should be different. The RpoS fusions were inserted at the endogenous locus, placing expression of the RpoS fusions under control of the endogenous *rpoS* promoter. The resulting *E. coli* strains were transformed with a plasmid containing the *sprE* promoter fused to an RFP (pDHL878) and the RFP levels were measured in stationary-phase cells with a plate reader. Since RpoS degradation ceases in stationary phase, all RpoS fusions should be present at approximately the same level if the tag is not interfering with protein synthesis or stability.

The *sprE* promoter is strongly expressed in stationary-phase cells and expression is highly RpoS-dependent. GFP fusions to the N-terminus and C-terminus of RpoS were essentially non-functional, except that the RpoS-mEGFP fusion showed some residual activity. I also constructed and evaluated an RpoS 'sandwich' fusion, in which GFP is inserted in-frame after amino acid 90 of RpoS. The choice of this insertion site was motived by sequence alignment of RpoS with the housekeeping sigma factor σ^{70} (data not shown). The sandwich fusions were non-functional (**Figure 4.8a**). Fusing the small 3xFLAG tag to the C-terminus of RpoS also resulted in a non-functional fusion (**Figure 4.8b**). These tagging difficulties are not too surprising since the sigma factor RpoS must bind to the RNA polymerase as well as to promoter sequences, and also interacts with various other proteins.

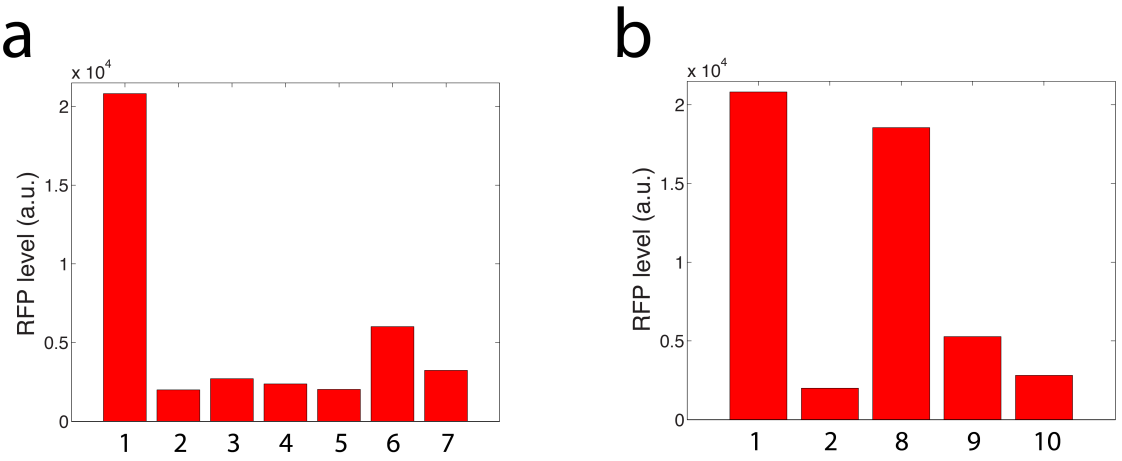


Figure 4.8. Traditional protein fusions to RpoS are non-functional. The activity of various RpoS fusions was assayed by measuring RFP expression from the *sprE* promoter (using plasmid pDHL878) in stationary-phase cells by plate reader. The *sprE* promoter is highly expressed in the wild type (1: DHL1178) and very weakly expressed in the Δ*rpoS* strain (2: DHL1179). (**a**) An N-terminal mEGFP fusion to RpoS is non-functional (3: DHL1180). RpoS sandwich fusions with mEGFP (4: DHL1181) and mGFPmut3 (5: DHL1182) are non-functional. The C-terminal fusion with mEGFP (6: DHL1185) shows some weak residual activity. The RpoS-mGFPmut3 fusion (7: DHL1186) is non-functional. (**b**) RpoS cannot tolerate the small 3xFLAG tag (~2 kDa). The *sprE* promoter is strongly expressed in a *sprE*-3x-FLAG *sspB*-3xFLAG background strain (8: DHL1189) but activity is greatly reduced if RpoS also has the 3xFLAG tag (9: DHL1190). The *sprE* promoter shows very weak expression in the Δ*rpoS sspB*-3xFLAG *sprE*-3xFLAG strain (10: DHL1191).

A recent study[250] adapted a clever approach[251] to tag a transcription factor that does not tolerate a traditional C- or N-terminal fluorescent protein tag. Hensel *et al.* fused a fluorescent protein (FP) followed by ubiquitin (Ub) to the N-terminus of the transcription factor (TF) in an *E. coli* strain that also co-expresses the yeast ubiquitin hydrolase Ubp1. The FP-Ub-TF fusion protein is then post-translationally cleaved by Ubp1 into the FP-Ub moiety and the free, non-tagged transcription factor. The number of transcription factors made in a certain time-interval can then be measured in real-time by quantifying the number of GFP molecules made. The GFP molecules were photo-bleached after each measurement, so that only newly made fluorescent molecules were measured. Since the GFP and TF were synthesized as a joint polypeptide chain, they also inherited the same

upstream randomness (coming from fluctuations in the transcription and translation machinery) and variability in the GFP signal reflects variability in the expression of the transcription factor. The GFP signal would for example contain much less information about the true transcription factor expression dynamics if a simple promoter fusion to the transcription factor's promoter would have been used instead.

I adapted this approach for tagging RpoS (**Figure 4.9**) and constructed two types of fusion proteins: mGFPmut3-Ub-RpoS and $RpoS_{1-32}$-mEGFP-Ub-RpoS.

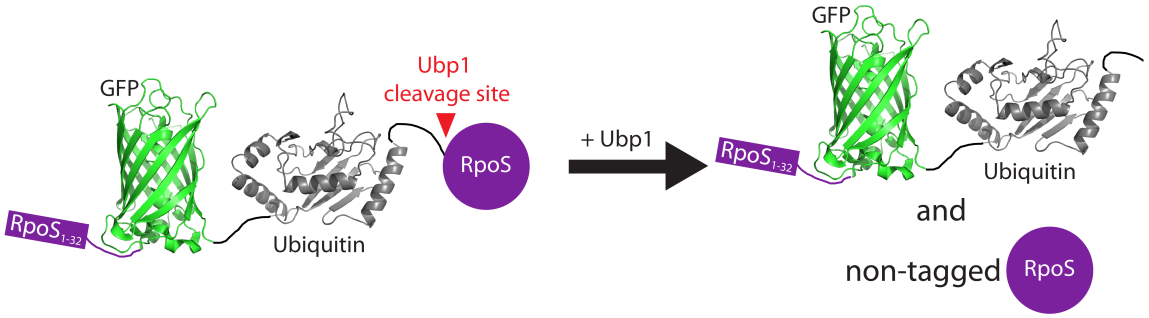


Figure 4.9. Schematic of the 'cotranslational activation by cleavage' (CoTrAC) technique[250] applied to RpoS. GFP and ubiquitin (Ub) are fused to the N-terminus of RpoS. Expression of yeast Ubp1 cleaves the fusion protein into a GFP-Ub moiety and non-tagged RpoS with a 1:1 stoichiometry. Two GFP-Ub-RpoS fusions with different N-termini were constructed since the N-terminus contains translational control elements and can strongly affect translational initiation because of differences in mRNA structure around the ribosome binding site[252]. One fusion has the RpoS N-terminus (corresponding to amino acids 1–32 of RpoS) and the second fusion has the mGFPmut3 N-terminus (this fusion is not shown in the cartoon).

The GFP-Ub-RpoS fusions were tested with the RpoS activity assay and, surprisingly, the fusions were fully functional (**Figure 4.10**).

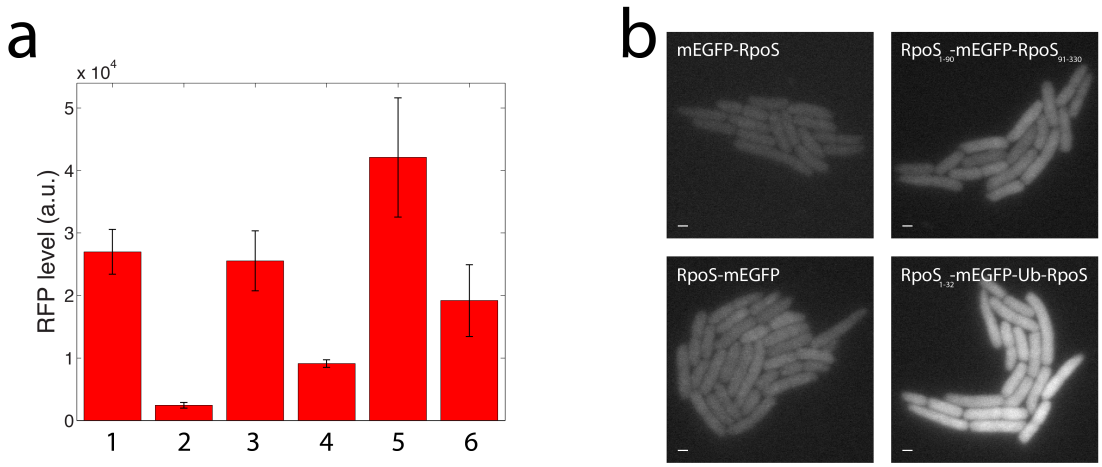


Figure 4.10. (**a**) The GFP-Ubiquitin (Ub) fusions to RpoS are active based on measuring RFP expression from the *sprE* promoter (using plasmid pDHL878) in stationary-phase cells. The *sprE* promoter is highly expressed in the wild type (1: DHL1178) and very weakly expressed in the Δ*rpoS* strain (2: DHL1179). Cells expressing the mGFPmut3-Ub-RpoS (3: DHL1184) or $RpoS_{1-32}$-mEGFP-Ub-RpoS (5: DHL1187) fusions produce RFP levels similar (or higher) than the wild type. Co-expression of Ubp1, which cleaves the fusion (see **Figure 4.9**), surprisingly reduces RpoS activity (see discussion in text). $RpoS_{1-32}$-mEGFP-Ub-RpoS in the presence of Ubp1 (6: DHL1188) is more active than mGFPmut3-Ub-RpoS with Ubp1 (4: DHL1202). Two to four biological replicates were measured of each strain. Error bars represent standard error. (**b**) Green fluorescence images of *E. coli* micro-colonies expressing the indicated constructs. The $RpoS_{1-32}$-mEGFP-Ub-RpoS fusion (w/o Ubp1, corresponding to 5 in a) exhibits the highest green fluorescence signal and is also the only functional fusion (out of the four fusions displayed in b). All images were subjected to the same grayscale scaling. Scale bars (white) are 1 μm.

The non-cleaved $RpoS_{1-32}$-mEGFP-Ub-RpoS fusion showed even a higher activity than the RpoS wild type in stationary phase (**Figure 4.10**, compare 5 to 1). This is likely due to an increase in the steady-state protein level of tagged RpoS compared to wild-type RpoS. This increase might be due to elevated synthesis (due to increased translation or higher mRNA levels) or more efficient protein folding. In fact, it has been reported that ubiquitin can dramatically increase the yield of ubiquitin fusion proteins in *E. coli* probably due to an inherent chaperone activity[253] that facilitates folding of the fusion. Further, the N- and C-termini of ubiquitin are located ~4.7 nm apart and at opposite ends

of the molecule. This creates a rigid spacer that might spatially separate the GFP and RpoS moieties during protein folding. Since the non-cleaved mGFPmut3-Ub-RpoS fusion showed similar activity to wild-type RpoS (**Figure 4.10**, compare 3 to 1) but lower activity than the non-cleaved $RpoS_{1-32}$-mEGFP-Ub-RpoS (**Figure 4.10**, compare 3 to 5), one could speculate that the RpoS N-terminus (i.e. $RpoS_{1-32}$) is responsible for the enhanced activity or that the difference is caused by the GFP. We have previously[205] observed that mGFPmut3 fusions can experience folding problems at 37 ˚C (the strains for the RpoS activity measurements were grown at 37 ˚C) because mGFPmut3 lacks the F64L folding mutation[129] (though this effect also depended on which protein was tagged). These two differences might explain why the effect of ubiquitin is only observed for the $RpoS_{1-32}$-mEGFP-Ub-RpoS fusion but not for the mGFPmut3-Ub-RpoS fusion.
Further, it is surprising that the activity of the GFP-ubiquitin-RpoS fusions decreases when the deubiquitination enzyme Ubp1 is expressed (**Figure 4.10**, compare 3 to 4 and 5 to 6). One could speculate that this is because Ubp1 cleavage leads to unfolding and degradation of some of the RpoS molecules but this type of effect has not been reported by Hensel *et al*.[250], though they tagged the λ repressor CI and not RpoS. Alternatively, the reduced signal could also be due to plasmid incompatibility between plasmids pDHL878 (pSC101 origin) and pJT184 (p15A origin). This hypothesis is currently tested. Despite the decrease in activity in the presence of Ubp1, the $RpoS_{1-32}$-mEGFP-Ub-RpoS fusion is still remarkably active, especially compared to the Δ*rpoS* strain (**Figure 4.10**, compare 6 to 2).

The green fluorescence signal of exponential-phase cells expressing the $RpoS_{1-32}$-mEGFP-Ub-RpoS fusion increased by ~50% when Ubp1 (plasmid pJT184) was co-expressed (**Figure 4.11**).

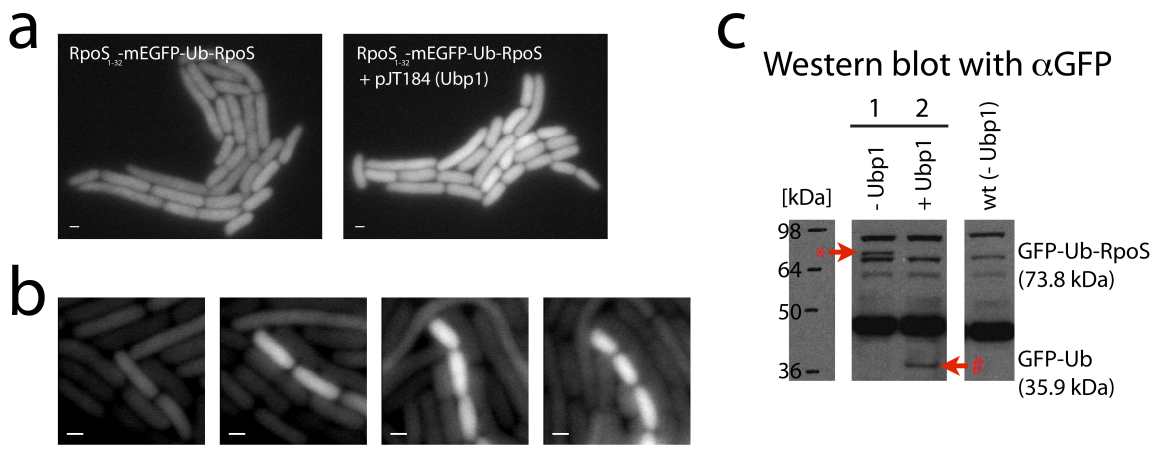


Figure 4.11. (**a**) Green fluorescence images of *E. coli* cells expressing the $RpoS_{1-32}$-mEGFP-Ub-RpoS fusion without (left, strain DHL1187) or with (right, strain DHL1188) Ubp1 (expressed from pJT184). Images were subjected to the same grayscale scaling. Scale bars (white) are 1 µm. (**b**) Time-series of cells expressing Ubp1 and the $RpoS_{1-32}$-mEGFP-Ub-RpoS fusion (strain DHL1188). One cell shows a sudden strong increase in the GFP signal and continues growing. The green fluorescence images correspond to time points 20 min apart. Images were subjected to the same grayscale scaling. Scale bars (white) are 1 µm. (**c**) Western blot with the αGFP antibody of cells expressing mEGFP-Ub-RpoS without (1: DHL1051) or with Ubp1 (2: DHL1050). Arrows highlight the GFP-Ub-RpoS (red *) and the GFP-Ub (red #) bands. Right lane (wild type without Ubp1) shows the non-specific bands of the αGFP antibody. Western with the αRpoS antibody was not done yet due to technical difficulties.

An increase in the green fluorescence signal is expected since RpoS is unstable in exponential-phase cells and GFP-tagged RpoS molecules are degraded before emitting any fluorescence (**Figure 4.11a**). Hence the $RpoS_{1-32}$-mEGFP-Ub-RpoS molecules are probably also degraded to some degree though very likely not as efficient as untagged RpoS since GFP can prevent or reduce active degradation (e.g. I observed that a RpoS750-mCherry fusion was even stable, data not shown). Co-expression of Ubp1 cleaves the fusion and separates the unstable RpoS from the GFP-Ub moiety. The GFP is then stable (not actively degraded) and can mature more efficiently, which should result in a higher GFP signal. The strain that harbors the $RpoS_{1-32}$-mEGFP-Ub-RpoS

fusions in the presence of Ubp1 has stable GFP whereas the strain that does not express Ubp1 contains the unstable GFP fusion protein.

4.3.3. The SprE-Venus fusion is hard to detect in exponential phase but the levels increase after exposure to a stress

In order to observe the RpoS-SprE feedback loop dynamics directly, the SprE protein was also tagged with a fluorescent protein. I tagged SprE at the C-terminus with the fluorescent protein Venus (**Figure 4.12a**). The Silhavy Lab (Princeton University) confirmed that the SprE-Venus fusion, which I constructed, is functional in the sense that RpoS is unstable in exponential-phase cells that express the SprE-Venus fusion protein (data not shown).

The SprE-Venus fusion is very weakly expressed in exponential-phase cells but can be detected by fluorescence microscopy, although the YFP levels are close to cellular autofluorescence (**Figure 4.12b**). The YFP level is also low in stationary-phase cells (data not shown). Interestingly, higher SprE-Venus expression levels and a substantial amount of cell-to-cell variability were observed after the cells were put on an agar pad, which lacked glucose, and starved on the agar pad for two hours (**Figure 4.12c**).

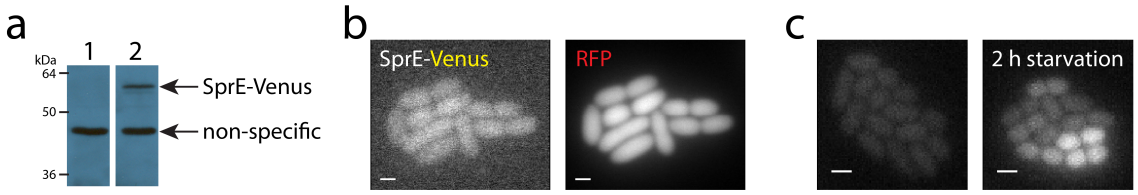


Figure 4.12. SprE-Venus is present in low numbers in exponential-phase cells. (**a**) Western blot with the αGFP antibody shows that stationary-phase cells with the SprE-Venus fusion (1: DHL399) have a band at the expected size of the SprE-Venus fusion protein (molecular weight = 64.3 kDa). Wild-type cells (2: MC4100) only have the non-specific band. The SprE-Venus band was barely detectable in exponential phase (data not shown). (**b**) The yellow fluorescence signal of exponential-phase cells with the SprE-Venus fusion (DHL813) is very low (left) but above cellular autofluorescence (data not shown). The cells also express a cytoplasmic RFP segmentation marker (right). Scale bars (white) are 1 μm. The YFP signal is also low in stationary-phase cells (data not shown). (**c**) Cells harboring a *sprE*-Venus allele (DHL352) were grown to late-exponential phase and spread on an agar pad that lacked glucose. The cells were imaged immediately after they were put on the agar pad (left image) and after ~2 hours (right image). Some cells showed a strong increase in the SprE-Venus level after the starvation. The two images were subjected to the same grayscale scaling. Scale bars (white) correspond to 1 μm.

The data suggests that SprE is present in very low numbers in exponential-phase cells, although SprE is of utmost importance for degrading RpoS since RpoS accumulates to high levels in exponential-phase cells that lack SprE (i.e. Δ*sprE* allele). This suggests that a very small number of SprE molecules are sufficient to control RpoS degradation in exponential phase. Since SprE is low abundant, one would expect relatively large fluctuations in the RpoS degradation rate and hence RpoS level. The data also indicates that SprE is increased when cells are first exposed to a stress (i.e. 2 h after starvation on agar pad without glucose) but not after the cells were exposed to the stress for a long time (e.g. stationary-phase cells do not have high levels of SprE-Venus). These findings are consistent with a model in which SprE is pulse-induced after exposure to a stress.

4.3.4. The *sprE* promoter is strongly RpoS-dependent in stationary-phase cells

Next, I set out to confirm the transcriptional arm of the RpoS-SprE negative feedback loop by using a transcriptional fusion between the *sprE* promoter and GFP. I analyzed GFP expression from the *sprE* promoter in stationary phase and found that stationary-phase cells taken from an overnight culture had very high GFP levels (**Figure 4.13**). This data suggests that the *sprE* promoter is either turned on in stationary phase or when cells transition into stationary phase. It is not possible to discriminate between these two scenarios based on this experiment because GFP is stable and only diluted due to cell growth, or potentially due to some protein turnover in stationary-phase cells.

I then analyzed expression from the *sprE* promoter in stationary-phase cells carrying a Δ*rpoS* deletion and found that the GFP signal was very low, showing that expression of the *sprE* promoter in stationary phase (or during the transition) strongly depends on the RpoS sigma factor. This result is in good agreement with previous reports using transcriptional fusions between LacZ and the *sprE* promoter[10,11].

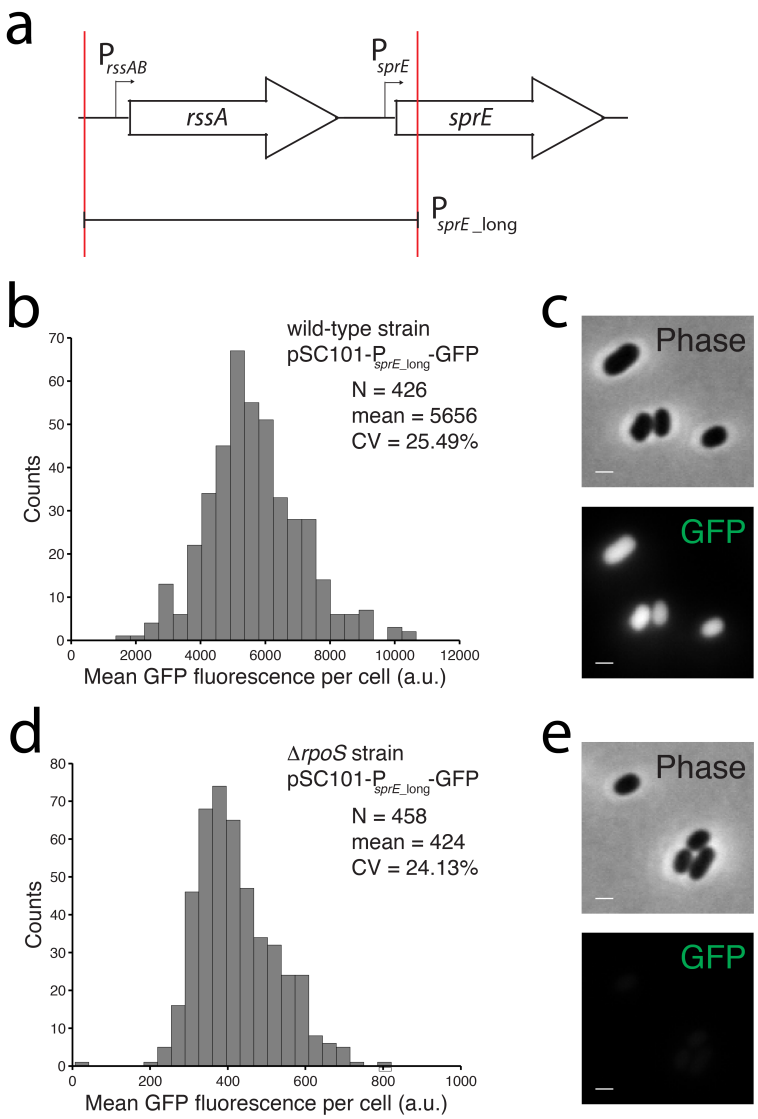


Figure 4.13. RpoS is necessary for transcription of the *sprE* promoter in stationary phase or during the transition into stationary phase. (**a**) Schematic of the *rssA* and *sprE* operon. P_{sprE_long} was used for the GFP promoter fusions since it contains all known *sprE* promoters. P_{rssA} and P_{sprE} were also tested; these promoters were weaker but also RpoS-dependent in stationary phase (data not shown). (**b**) Expression of the *sprE* promoter in stationary-phase wild-type cells (grown overnight in liquid culture, strain DHL319), (**c**) phase and green fluorescence images of representative cells. (**d**) No expression of the *sprE* promoter was observed in stationary-phase cells (overnight culture) with the Δ*rpoS* allele (strain DHL320), (**e**) phase and green fluorescence images of representative cells. Scale bars (white) are 1 μm.

4.3.5. Characterization of the RpoS and SprE protein levels by Western blot

I tried to characterize the RpoS-SprE negative feedback loop with bulk measurements to confirm the protein-protein interaction arm of the feedback loop and other published results, and also to measure the half-life of SprE in exponential and stationary phase.

First, I measured the exponential-phase RpoS level in the wild type and in a Δ*sprE* strain. My Western blot analysis with a commercially available monoclonal αRpoS antibody showed that SprE is necessary for RpoS degradation in exponential phase (**Figure 4.14**) and hence confirmed the protein-protein interaction arm of the negative feedback loop. This was unsurprising since it is a well established result in the field[10,11,229,230,235].

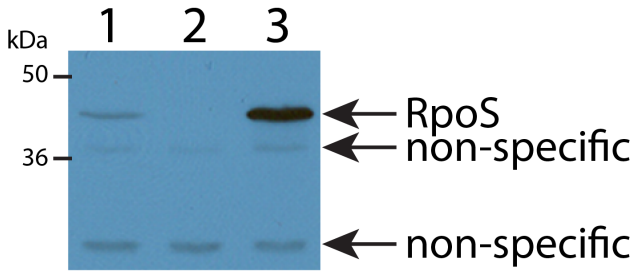


Figure 4.14. SprE is necessary for RpoS degradation in exponential-phase cells (OD_{600} = 0.2–0.3). Western blot with the αRpoS antibody showed that the RpoS level is low in the wild type (1: MC4100) but high in the Δ*sprE* strain (3: DHL241). This result confirmed that SprE is necessary for RpoS degradation in exponential phase[10,11,229,230,235]. No RpoS band was detected in the Δ*rpoS* strain (2: DHL51). The molecular weight of RpoS is 38.0 kDa.

Unfortunately, there are no commercially available antibodies against SprE. Although I received an αSprE antibody from the Silhavy lab, I did not find this antibody to work effectively for Western blotting (data not shown).

To get around this problem, I decided to attach the small 3xFLAG tag to the C-terminus of SprE by modifying the endogenous *sprE* gene. The 3xFLAG tag is only 2 kDa (i.e.

less than 10% the size of GFP) and antibodies with high sensitivity and specificity are commercially available against the FLAG tag (e.g. A8592, clone M2, Sigma-Aldrich). Using this antibody I showed that the SprE-3xFLAG fusion protein degrades RpoS in exponential-phase cells identically to SprE without the FLAG tag (**Figure 4.15**, compare 1 to 4).

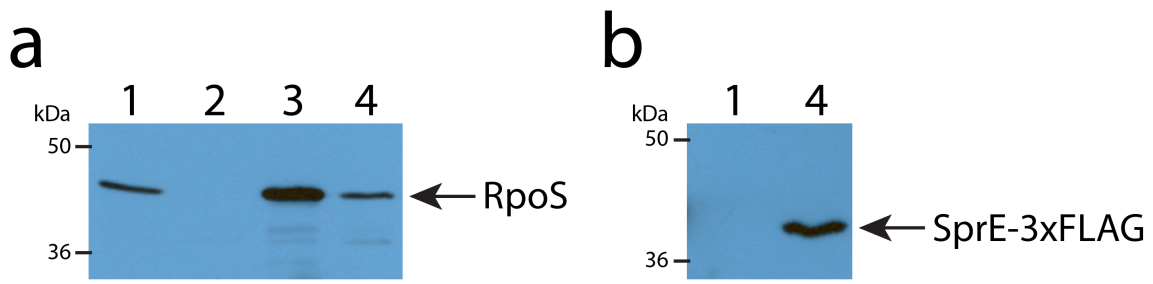


Figure 4.15. SprE-3xFLAG behaves identical to untagged SprE and degrades RpoS in exponential-phase cells (OD_{600} = 0.3–0.4). (**a**) Western blot with the αRpoS antibody showed that RpoS is low in the wild type (1: MC4100) and in the *sprE*-3xFLAG tag strain (4: DHL290), The RpoS level strongly increased in the Δ*sprE* strain (3: DHL241) and no RpoS band was observed in the Δ*rpoS* strain (2: DHL51). (**b**) Western blot with the αFLAG antibody resulted in a strong SprE-3xFLAG tag band for the strain that has the endogenous *sprE* gene tagged with the 3xFLAG tag (4: DHL290). No band was observed for the wild-type strain (1: MC4100), which has untagged *sprE*. The molecular weight of the SprE-3xFLAG tag fusion protein is 39.3 kDa.

SprE has been reported to be undetectable in early exponential phase (below OD_{600} = 0.5) and very low in mid-exponential phase (OD_{600} = 0.9)[16]. This observation is in agreement with my microscopy data of the SprE-Venus fusion (strain DHL812), which resulted in a yellow fluorescence signal that is only slightly above cellular autofluorescence in exponential-phase cells (**Figure 4.12b**). I observed a rather strong signal for the SprE-3xFLAG fusion protein in exponential-phase cells, but this is likely due to the high sensitivity of the αFLAG antibody. It is not possible to infer the number of

SprE molecules per cell from the band intensity alone without doing further experiments (e.g. quantitative Western blot with purified SprE protein as a standard).

I then analyzed by Western blotting how the SprE-3xFLAG level depends on RpoS. Surprisingly, I found that the SprE-3xFLAG protein level is independent of RpoS in exponential-phase cells (**Figure 4.16**) and only weakly dependent on RpoS in stationary-phase cells (**Figure 4.17**).

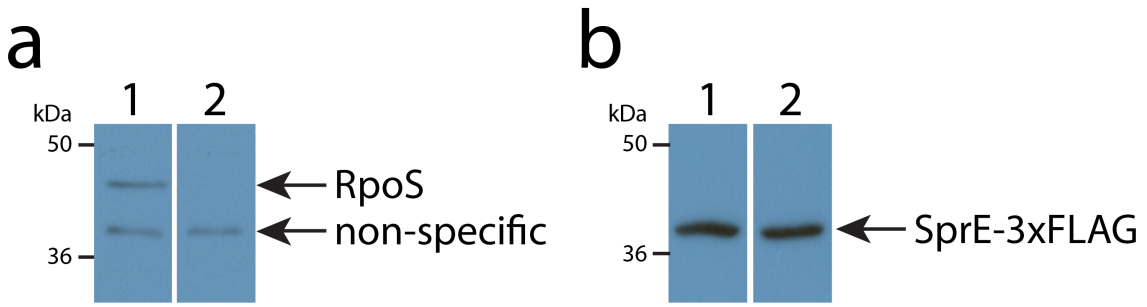


Figure 4.16. The SprE protein level is independent of RpoS in exponential-phase cells (OD_{600} = 0.3). (**a**) Western blot with the αRpoS antibody. Both strains (lanes 1 and 2) harbor the *sprE*-3xFLAG tag allele; the first strain (1: DHL307) has the *rpoS* wild-type allele whereas the second strain has the Δ*rpoS* allele (2: DHL309). The band around 40 kDa is non-specific. (**b**) Western blot with the αFLAG antibody. Both strains expressed similar levels of the SprE-3xFLAG tag fusion despite the extreme difference in RpoS. The first strain has wild-type *rpoS* (1: DHL307) whereas the second strain has the Δ*rpoS* allele (2: DHL309). This result shows that RpoS is not affecting the SprE level in exponential-phase cells and suggests instead that SprE is made exclusively from a σ^{70}-dependent promoter in exponential phase.

In stationary phase, I found that the SprE-3xFLAG level is twofold decreased in the Δ*rpoS* strain compared to the wild-type strain (**Figure 4.17**).

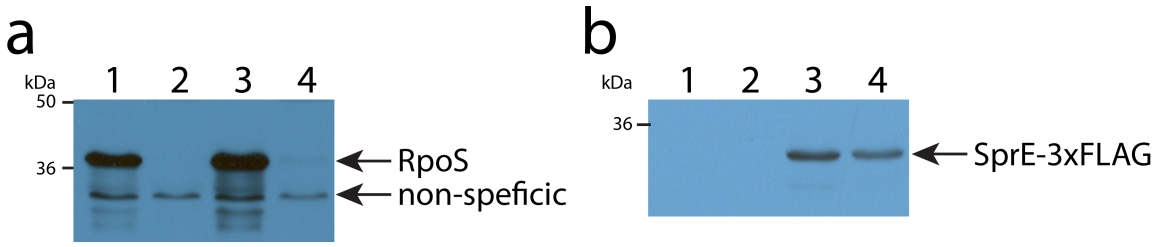


Figure 4.17. The level of the SprE-3xFLAG tag fusion protein is weakly RpoS-dependent in stationary-phase cells (overnight culture). (**a**) Western blot with the αRpoS antibody showed that RpoS strongly accumulates in wild-type stationary-phase cells (1: MC4100) and in stationary-phase cells with the *sprE*-3xFLAG allele (3: DHL290). The RpoS band was absent in the Δ*rpoS* strain (2: DHL51) and the Δ*rpoS sprE*-3xFLAG tag strain (4: DHL295). (**b**) Western blot with the αFLAG antibody. Only the strains of the third (3: DHL290) and fourth (4: DHL295) lane have the *sprE*-3xFLAG tag allele and resulted in a band. The SprE-3xFLAG tag fusion protein level was ~50% lower in the Δ*rpoS* strain (4: DHL295) compared to the wild type (3: DHL290).

No previous study has measured the endogenous SprE protein level in exponential-phase cells in the presence and absence of RpoS to determine how RpoS affects the SprE protein level in exponential phase. This is likely because the αSprE antibody used was not sensitive enough; the αSprE antibody can only detect SprE in late-exponential phase[16] or in stationary phase[11], or when SprE[11] or RpoS[10] were over-expressed from a plasmid.

The SprE-3xFLAG tag protein level could be measured in exponential-phase cells without overexpression, but that required attaching the 3xFLAG tag to the C-terminus of the SprE protein for detection purposes. It is possible that the 3xFLAG tag is intrusive and changed the regulation of SprE, although it is unlikely that the 3xFLAG tag affects regulation of SprE synthesis and the sigma factor dependence of the promoter (e.g. due to mutations that rewired the transcriptional regulatory network). Further, I showed that the SprE-3xFLAG tag protein is functional and degrades RpoS in exponential-phase cells. The SprE-3xFLAG tag strain was rebuilt with two different methods and I obtained the same result (data not shown), indicating that the data obtained with the SprE-

3xFLAG tag strain is likely not due to an artifact introduced during strain construction. It will be interesting to repeat the same type of Western blot analysis using my SprE-Venus fusion as another independent validation. My result is quite surprising and strongly argues that the transcriptional arm of the RpoS-SprE feedback loop is not active in early exponential phase since SprE is made independent or RpoS.

Two studies have previously analyzed how the stationary-phase SprE protein level depends on RpoS. Both studies used Western blotting with the αSprE antibody (from the Silhavy Lab) to detect the endogenous SprE protein. One study found that SprE was not made at all in stationary-phase cells in the absence of RpoS[16] whereas the second study observed a much smaller RpoS dependency[11]. Ruiz *et al.* found that the SprE protein level depended on RpoS in stationary-phase cells (OD_{600} = ~3.0) but they have also reported that some SprE protein was made in the absence of RpoS[11], which shows that SprE is also made from an σ^{70}-regulated promoter, independent of RpoS. This suggests that the *sprE* promoter is transcribed by both RpoS and σ^{70} and that the relative sigma factor contribution depends on the exact state of the cells (i.e. exponential phase or stationary phase). The observation by Ruiz *et al*.[11] agrees more closely with my Western blot result (**Figure 4.17b**). The remaining differences could be due to sample preparation and the fact that their stationary-phase cells were harvested at OD_{600} = ~3.0 whereas I used an overnight culture.

Overall, the data is consistent with a model in which SprE is made independently of RpoS in exponential phase, when RpoS is low, but the SprE synthesis during transition into stationary phase, when RpoS stabilizes, depends on RpoS. The RpoS-independent synthesis of SprE in exponential phase is sufficient for fast degradation of RpoS. The negative feedback loop is thus broken in exponential phase (i.e. SprE synthesis is independent of RpoS) and in stationary phase (i.e. RpoS is not degraded) but operates

when cells transition into stationary phase. This model needs to be further tested by experiments.

4.3.6. Transition of *E. coli* cells into stationary phase and the induction of RpoS target genes can be studied with the mother machine

The second hypotheses that was generated by the mathematical modeling, suggests that another potential role of the lagged feedback loop response is to allow a fast induction of stress response target genes. I set out to study this experimentally by analyzing RpoS-regulated promoters in the wild-type background and compare the target gene response to feedback loop mutants (in which e.g. *sprE* transcription is constitutive).

It is difficult to study how *E. coli* cells transition from exponential growth into stationary phase and induce stress response genes when grown on an agar pad. *E. coli* cells growing into micro-colonies on agar pads do not stop growing after reaching a single monolayer and instead grow into a second layer. Even at that point most cells do not enter stationary phase likely because the agar pad provides fresh nutrients and extra-cellular signaling molecules diffuse away before they reach a high enough concentration. However, based on measuring cell size and the expression of a fluorescent reporter, I occasionally observed that a very small number of cells were entering stationary phase (**Figure 4.18**).

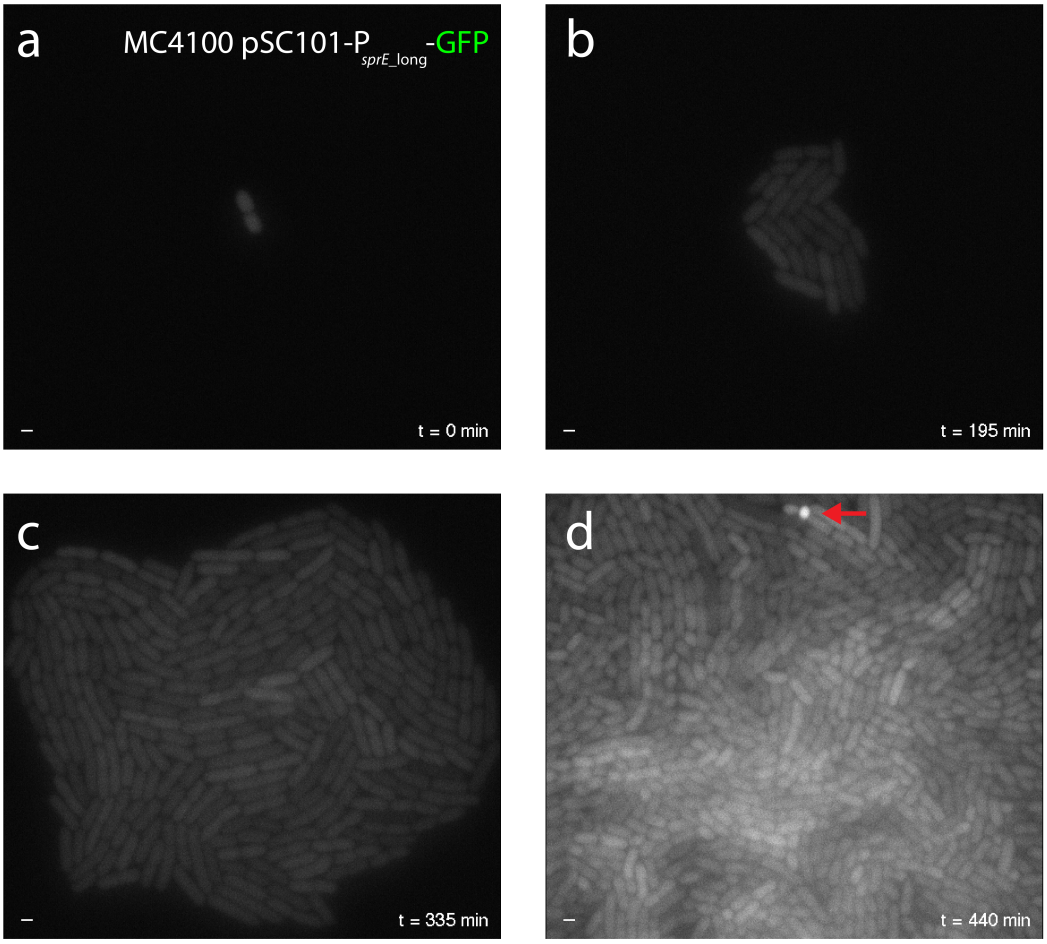


Figure 4.18. The transition of single *E. coli* cells from exponential growth into stationary phase is difficult to study when the cells are grown on an agar pad. (**a**, **b**) The growth of a single cell into a micro-colony can be followed by time-lapse microcopy. (**c**) The cells grow into a monolayer that fills up the field of view of the camera and then start to 'pile up' and grow in a second layer or even perpendicular to the imaging surface. (**d**) Occasionally single cells seemed to enter stationary phase (see bright cell with red arrow) but this happens very infrequently. The strain carries a pSC101 plasmid with the *sprE* promoter GFP fusion (strain DHL319). Stationary-phase cells that were grown in liquid culture are small in size, have a round shape and strongly express the *sprE* promoter GFP fusion (see **Figure 4.13c** for stationary-phase cells from an overnight liquid culture). Timestamps are in lower right corner of the images. Scale bars (white) are 1 µm.

Agar pads are not well suited to study the transition of *E. coli* cells into stationary phase and the induction of stress response genes, because (i) not many cells enter stationary phase, (ii) the growth medium is hard to exchange to starve or stress the cells and (iii)

the agar pad introduces cell-to-cell variability likely due to spatial inhomogeneities (Nate Lord, personal communication). Further, many cells are in close proximity in a micro-colony (see **Figure 4.18c** or **d**). This makes quantitative imaging difficult because neighboring cells contribute to the fluorescence intensity of the cell studied, due to the width of the point spread function (Nate Lord, personal communication).

Fortunately these problems were overcome by using a microfluidic device with growth channels, colloquially called the 'mother machine'[254]. This device allows quantitative imaging of cell lineages in isolation in an unperturbed environment. The growth medium can also be exchanged relatively quickly to starve or stress the cells. I used the mother machine in collaboration with Nate Lord (PhD student, Paulsson Lab) to follow how single *E. coli* cells transition from exponential phase into stationary phase. The cells were grown in M9 medium with 10% (v/v) LB for a few hours. The growth medium was then exchanged against 1x PBS to starve the cells and force entry into stationary phase. The strain that was imaged (DHL452) has an *rpoS* promoter GFP fusion integrated in the chromosome and carries a plasmid with a *sprE* promoter mCherry RFP fusion. We observed that the cells strongly expressed the two transcriptional reporters after the medium was exchanged against 1x PBS (**Figure 4.19** and **Video 4.1**).

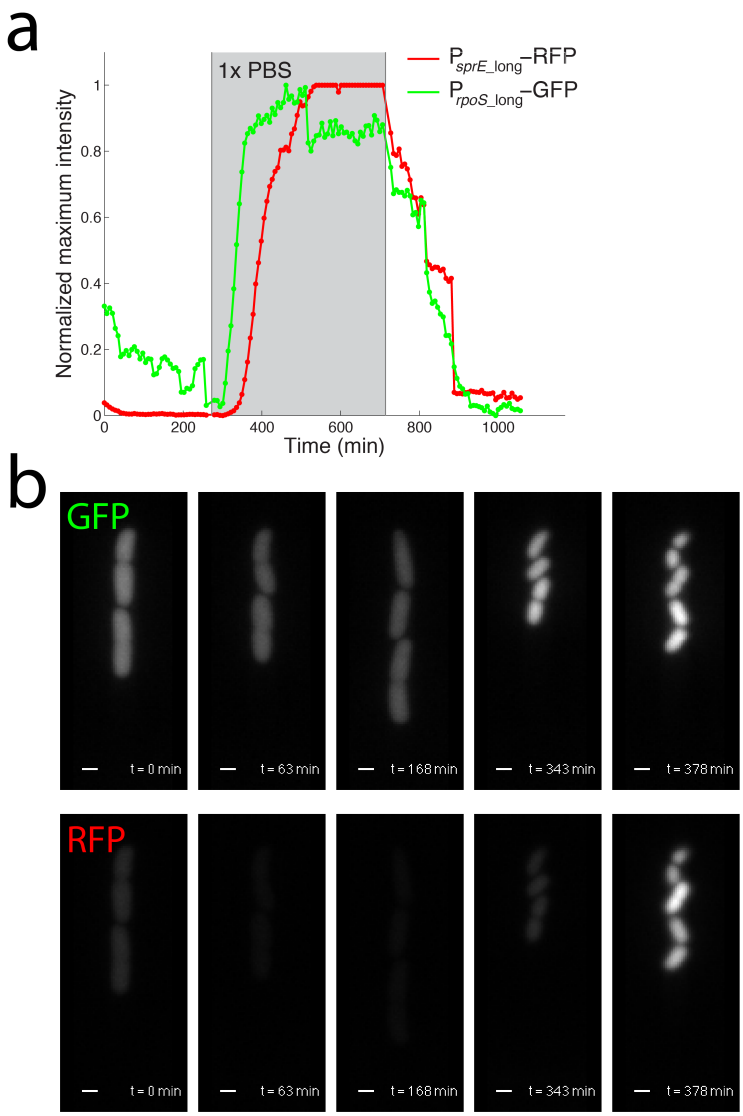


Figure 4.19. Transition of single *E. coli* cells into stationary phase can be observed by using the mother machine. (**a**) The *E. coli* cells were loaded into the mother machine and grown in a rich medium for ~4 h. The growth medium was then exchanged against 1x PBS (gray box in graph) to starve the cells and force entry into stationary phase. The graph shows how the *rpoS* and *sprE* promoter turn on after the media switch. The normalized maximum fluorescence intensity in the channel is plotted. Both promoters show a fast induction. The RFP signal is temporary truncated (from 550–700 min) because of camera saturation. The cells were released from the starvation after ~12 h. (**b**) Green fluorescence (upper) and red fluorescence (lower) images are shown for respective time points before and after the media switch. The cell size decreased dramatically after the media switch and the cells also adopted the typical round shape of stationary-phase cells (which is even more pronounced later in the experiment, see **Video 4.1**). Scale bars (white) are 1 µm.

Interestingly, the RFP signal of the *sprE* promoter fusion seems to increase almost as fast as the GFP signal of the *rpoS* promoter fusion after the cells were exposed to the stress. A direct comparison between the two signals is difficult because different fluorescent proteins are used and GFP matures quickly (6.5 min)[255] and with a single exponential[129] whereas mCherry matures slowly (~45min)[115] and with a double exponential[161]. Since mCherry matures much slower than GFP, it is possible that the *sprE* promoter is actually turning on faster than the *rpoS* promoter, consistent with the mathematical model. Further experiments will be necessary to show this effect more clearly and also that the fast target gene response depends on the RpoS-SprE feedback loop dynamics.

Interestingly, the *rpoS* promoter was strongly pulse-induced when cells were exposed to a stress but *rpoS* transcription ceased after the cells adapted to the stress (see **Figure 4.19a**, no increase of green line after 400 min).

I also obtained preliminary data (data not shown) suggesting that expression of the *sprE* promoter depends on how rapidly conditions change. For example, we observed that cells close to the feeding channel of the mother machine showed a much stronger response when the starvation media was introduced than cells further away from the feeding channel, though all cells where in the end immersed in the same media, and the cells in different positions have identical growth rates in exponential phase experiments. Transiently, however, the cells closer to the feeding channel would experience a more rapid change than the cells that were located further away from the feeding channel. This effect is interesting physiologically, as it may suggest an adaptive strategy to stress, but must be further investigated.

The mother machine provides an ideal setup to study the induction of RpoS target genes during entry into stationary phase and after exposure to a stress. I selected a handful of supposedly RpoS-regulated promoters (i.e. P_{rssAB}, P_{sprE_long}, P_{wrbA} and P_{bolA}) and

confirmed that these promoters are strongly RpoS-dependent by measuring GFP expression of the respective promoter fusions in the wild-type and the Δ*rpoS* strain (data not shown). I also choose three control promoters (P_{lacZ}, P_{rpoD} and P_{rpsU}), the expression of which I confirmed to be independent of RpoS (data not shown). I then constructed *E. coli* strains with the promoter fusions integrated as single copies at the *att*Tn7 site in the *E. coli* chromosome (strains DHL1205–1209 and DHL1211–1213). It will be interesting to analyze these transcriptional reporters in the wild type and in various feedback loop mutants (e.g. where expression of *sprE* is exclusively under control of σ^{70}) to analyze the effect of the RpoS-SprE feedback loop on the RpoS target gene response in greater depth.

4.4. Discussion

I presented in this chapter progress towards understanding the functional role of the RpoS-SprE negative feedback loop in *E. coli* cells under non-stressed conditions (like exponential phase) and after exposure to a stress. The mathematical modeling focused on analyzing the potential effects of feedback on fluctuations and the speed of response to environmental perturbations. Using measured RpoS and SprE parameters in the fluctuation analysis showed that this feedback loop should not be able suppress spontaneous fluctuations, because of noisy signaling, long lags in the loop, and weakly cooperative feedback effects. Only fluctuations from slowly changing environmental changes could be suppressed by such systems, and even then the effects are expected to be rather limited. The adjustment rate analysis by contrast showed that the feedback loop seems perfectly designed to minimize the response time in RpoS-regulated genes.

No functional fusions to RpoS have been reported in the literature. For an initial estimate I therefore constructed a fluorescent protein-based RpoS degradation reporter and confirmed that degradation of this reporter depends on SprE. I also found that a small fraction (~0.1%) of exponential-phase cells contained much higher levels of the RpoS degradation reporter than the majority of cells. This subpopulation could be more stress resistant and may insure against complete population collapse in the event of sudden stress, e.g. nutritional downshifts or antibiotic exposure[5]. It still needs to be verified that these cells are functionally different and for example are more likely to survive stressful conditions.

I was also eventually able to obtain functional GFP fusions to RpoS by using ubiquitin as a spacer, though the fusions need to be further verified to know for sure that there is no interference. The GFP-Ub-RpoS fusions can also be used in combination with an

enzyme that cleaves the fusion co-translationally to quantify the number of RpoS molecules made during a certain time interval. I further constructed a functional fusion to SprE with the yellow fluorescent protein Venus. The levels of the SprE-Venus fusion were very low in exponential-phase and stationary-phase cells and hard to detect by epi-fluorescence microscopy. A starvation experiment showed that the SprE-Venus fusion accumulated in some cells after the cells were exposed to a stress (i.e. lack of glucose). Using the same tagging strategy, but replacing the fluorescent proteins with the smaller SNAP and/or HaloTag, which I have demonstrated to have single-molecule sensitivity (partially presented in Chapter 2), I hope to obtain an assay to measure RpoS and SprE levels even in exponential phase, where levels are extremely low and half-lives are too short for fluorophore maturation. Such tags would be the basis for a more systematic study of RpoS fluctuations, hopefully also in conjunction with the mRNA counting methods presented in Chapter 3.

I further analyzed whether the suggested feedback loop is effective in different growth phases. First I confirmed the already established result that RpoS degradation in exponential phase is indeed SprE dependent. Second I confirmed that expression of the *sprE* promoter, in stationary-phase cells or during the transition into stationary phase, is indeed highly RpoS-dependent. I then analyzed how SprE levels depend on RpoS, and tagged SprE with the 3xFLAG tag to visualize it by Western blotting. Analysis revealed that the SprE protein level was independent of RpoS in exponential-phase cells but two-fold higher in stationary phase. This observation is consistent with a leaky gene expression model in which SprE is made at a low rate from a σ^{70}-regulated promoter in exponential phase, where RpoS levels are extremely low, but pulse-induced in an RpoS-dependent manner during the transition into stationary phase. This suggests that the RpoS-SprE feedback loop only kicks in towards the end of exponential phase growth. Preliminary mathematical analysis suggests that this could further support the

conclusion that the feedback loop minimizes the delay in stress response: turning on the feedback loop towards the end of the adjustment phase allows the system to initially respond with the high rate of a system adjusting towards a much elevated steady state, but then stop production at a more moderate level. It could thus produce a similar dynamic as an incoherent feed-forward loop, where a direct positive interaction promotes initial high speed, and a delayed negative interaction ensures that final levels do not increase too much.

Analysis of transcriptional fusions to the *sprE* and *rpoS* promoters indeed revealed that both promoters are quickly induced when cells transition into stationary phase after nutrient starvation and both promoters also seemed to exhibit an induction pulse.

In summary, an initial fluctuation analysis suggests the existence of a small sub-population of exponentially growing cells with strongly elevated RpoS levels, which could be better prepared to deal with future stresses. Furthermore, the successful construction of functional RpoS reporter fusions will allow us to more carefully analyze this system in individual cells. However, regardless of whether the feedback loop suppresses or amplifies fluctuations, the mathematical and experimental results also strongly suggest that the RpoS system has many unusual features that all contribute to an extremely fast response in RpoS controlled genes: extremely short half-lives, negative feedback control, and long relative lags in the feedback loop. My results also suggest that the negative feedback loop only kicks in later in the adjustment phase, which could speed up adjustments even further.

4.5. Materials and Methods

Plasmid constructions

All plasmids used in this study are listed in **Table 4.1**.

Table 4.1

Plasmid	Description	Antibiotic marker	Reference
pDHL16	pUC19-P_{rpoS}-*rpoS750*	Amp	This study
pDHL17	pUC19-Venus-T1 terminator	Amp	This study
pDHL19	pUC19-FRT Kan FRT	Amp, Kan	[205]
pDHL20	pET15b-6xHis-RpoS	Amp	Lab collection
pDHL23	pUC19-P_{rpoS}-*rpoS750*-Venus-T1 terminator	Amp	This study
pDHL39	pUC19-P_{rpoS}-*rpoS750*-Venus-T1 terminator -FRT Kan FRT	Amp, Kan	This study
pDHL138	pSC101-P_{rpoS_long}-GFPmut2-T1 terminator	Kan	This study
pDHL229	pUC19-3xFLAG-FRT Kan FRT	Amp, Kan	This study
pDHL304	pSC101-P_{sprE_long}-GFPmut2-T1 terminator	Kan	This study
pDHL374	pSC101-P_{LlacO1}-mCherry-T1 terminator	Amp	Lab collection
pDHL391	pSC101-P_{sprE_long}-GFPmut2-T1 terminator	Amp	This study
pDHL450	pSC101-P_{sprE_long}-mCherry-T1 terminator	Amp	This study
pDHL580	pUC19-linker-mGFPmut3-FRT Kan FRT	Amp, Kan	[205]
pDHL586	pSC101-P_{A1O4}-mKate2-*ssrA*(LAA)-T1 terminator-*lacI*^{q}	Amp	Lab collection
pDHL819	pSC101-P_{sprE_long}- mKate2-T1 terminator	Amp	This study
pDHL876	pUC19-*tsr*-mGFPmut3-Ub-*rpoS*-FRT Kan FRT	Amp, Kan	This study
pDHL878	pSC101-P_{sprE_long}-hybrid mKate2-T1 terminator	Amp	This study
pDHL896	pUC19-$rpoS_{(1-90)}$-mGFPmut3-$rpoS_{(91-330)}$-FRT Kan FRT	Amp, Kan	This study
pDHL919	pUC19-mGFPmut3-*rpoS*-FRT Kan FRT	Amp, Kan	This study
pDHL962	pUC19-linker-mEGFP-FRT Kan FRT	Amp, Kan	This study
pDHL1014	pUC19-*tsr*-mEGFP-Ub-*rpoS*-FRT Kan FRT	Amp, Kan	This study
pDHL1024	pUC19-$rpoS_{(1-90)}$-mEGFP-$rpoS_{(91-330)}$-FRT Kan FRT	Amp, Kan	This study
pDHL1025	pUC19-mEGFP-*rpoS*-FRT Kan FRT	Amp, Kan	This study
pDHL1100	pSC101-P_{LlacO1}-*sprE*-3xFLAG-T1 terminator	Amp	This study
pNDL194	pUC19-P_{RNAI}-hybrid mKate2-T1 terminator-FRT Kan FRT	Amp, Kan	Nate Lord (unpublished)
pUC19	High copy number cloning vector	Amp	Invitrogen
pPM16	pSC101-P_{LlacO1}-Venus-T1 terminator	Amp	Lab collection
pJT184	Ubp1 expression vector, p15A origin	Cm	[251]

Plasmid construction was verified with analytical restriction digest. Restriction enzymes were purchased from NEB. The cloning steps that involved PCR where verified by DNA sequencing. Phusion (Finnzymes) and Vent (NEB) polymerases were used for PCR amplification.

pDHL16 was built by the amplifying the full-length *rpoS* promoter (P_{rpoS_long}) and the first 750 base pairs of the *rpoS* open-reading frame (*rpoS*750) from genomic DNA with primers DHL_P13_F and DHL_P14_R. The full-length *rpoS* promoter includes the preceding *nlpD* gene and its promoter sequences. The PCR product was gel-purified, digested with EcoRI and SacI and ligated in pUC19, which was cut with the same restriction enzymes.

pDHL17 was built by amplifying Venus and the T1 terminator from pPM1 (courtesy of Per Malkus) with primers DHL_P15_F and DHL_P16_R. The PCR was gel purified, digested with SacI and XmaI and ligated into pUC19 that was also digested with SacI and XmaI.

pDHL23 was built by digesting pDHL17 with SacI and XmaI. The Venus-T1 terminator fragment was gel purified and subcloned into pDHL16, which was previously digested with SacI and XmaI.

pDHL39 was built by digesting pDHL23 with EcoRI and XmaI and gel purifying the P_{rpoS}-*rpoS*750-Venus-T1 terminator fragment, which was then ligated into EcoRI/XmaI-cut pDHL19.

pDHL138 was built by PCR amplifying the full-length *rpoS* promoter (P_{rpoS_long}), which includes the *nlpD* gene and sequences upstream of *nlpD*, from genomic DNA with primers DHL_P55_F and DHL_P57_R. The PCR product was digested with BamHI and XhoI, gel-purified and ligated into pUA139, which was previously digested with the enzymes BamHI and XhoI. pUA139 is one of the two parent vectors of the Uri Alon *E. coli* promoter library[256]. The *rpoS* promoter of the Uri Alon library corresponds to the

intergenic region between the *rpoS* and *nlpD* gene though the main *rpoS* promoter is positioned in the coding region of the *nlpD* gene. The region upstream of *nlpD* and the intergenic region between *nlpD* and *rpoS* also contain promoters for *rpoS* transcription. Promoter annotation was obtained from EcoCyc (http://www.ecocyc.com/).

pDHL229 was built by PCR amplifying the 3xFLAG tag from plasmid pSUB11[257] with primers DHL_P105_F and DHL_P106_R. The PCR fragment was then cut with SacI and XmaI, gel purified and ligated into pDHL19, which was digested with SacI and XmaI.

pDHL304 was built by PCR amplifying the full-length *sprE* promoter (P_{sprE_long}) from genomic DNA with primers DHL_P118_F and DHL_P119_R. The PCR product was digested with BamHI and XhoI, gel-purified and ligated into pUA139[256], which was previously digested with the enzymes BamHI and XhoI. The full-length *sprE* promoter contains the sequences upstream of *sprE* (*rssB*) including the *rssA* gene since *sprE* is the second gene of a bicistronic operon (*rssAB*).

pDHL391 was built by amplifying the Amp resistance marker from pUC19 with primers DHL_P161_F and DHL_P162_R. The resulting PCR fragment was then digested with SacI and BglII and ligated into pDHL304, which was also digest with SacI and BglII.

pDHL450 was built by amplifying mCherry and the T1 terminator from pDHL374 with primers DHL_P163_F and DHL_P172_R. The PCR fragment was digest with XhoI and AvrII, gel-purified and ligated intro pDHL391, which was cut with the same restriction enzymes.

pDHL819 was built by PCR amplifying mKate2 from pDHL586 with primers DHL_P404_F and DHL_P405_R. The resulting PCR fragment was then digested with XhoI and SbfI and ligated into pDHL450, which was cut with XhoI and SbfI.

pDHL876 was built by isothermal assembly[181]. Four PCR fragments were ligated into pDHL19, which was digested with EcoRI and XmaI. PCR fragment one corresponds to the membrane-targeting domain of Tsr[191], which was amplified from pSV152Tsr[191] with

primes DHL_P433_F and DHL_P434_R. Fragment two corresponds to mGFPmut3 and was amplified from pDHL580 with primers DHL_P435_F and DHL_P436_R. The third fragment corresponds to yeast ubiquitin and was amplified from pUB23-M-βgal[251] using primers DHL_P437_F and DHL_P438_R. The fourth fragment corresponds to the *rpoS* coding region and was PCR amplified from genomic DNA with primers DHL_P439_F and DHL_P442_R.

pDHL878 was built by inserting two oligo sites into pDHL819 to exchange N-terminus of mKate2 (i.e. MVSE) against the N-terminus of mCherry (i.e. MVSKGEENNMA). The two sites were made by annealing oligo DHL_P487_F with DHL_P488_R and DHL_P489_F with DHL_P490_R. I found out that the N-terminus of mCherry is beneficial for high-level production of the mCherry fluorescent protein likely because it contains a second ribosome-binding site (aaggag), which increases translation. Adding the N-terminus to mKate2 also increased expression levels of mKate2. The resulting mKate2 variant was named 'hybrid mKate2'.

pDHL896 was built by isothermal assembly[181]. Three inserts were ligated into SacI/XmaI-digested pDHL19. The first PCR fragment corresponds to the first 90 amino acids of RpoS (i.e. $RpoS_{(1-90)}$) and was PCR amplified from pDHL20 with primers DHL_P466_F and DHL_P467_R. The second fragment corresponds to mGFPmut3 and was amplified from pDHL580 with primers DHL_P468_F and DHL_P469_R. The third fragment corresponds to the last 240 amino acids of RpoS (i.e. $RpoS_{(91-330)}$) and was PCR amplified from pDHL20 with primers DHL_P470_F and DHL_P442_R.

pDHL919 was built by isothermal assembly. Two PCR fragments were ligated into SacI/XmaI-digested pDHL19. The first fragment corresponds to mGFPmut3 and was PCR amplified from pDHL580 with primers DHL_P482_F and DHL_P483_R. The second fragment corresponds to the RpoS and was PCR amplified from pDHL20 with primers DHL_P484_F and DHL_P442_R.

pDHL962 was built by amplifying mEGFP from pGEMHE-XfA4-mEGFP[126] with primers DHL_P295_F and DHL_P543_R. The resulting PCR fragment was digested with SacI and XmaI, gel purified and ligated into pDHL19, which was also cut with SacI and XmaI.

pDHL1014 was built by combining three PCR fragments with isothermal assembly[181]. The first fragment, which corresponds to the pUC19 origin of replication, the *bla* gene and the Tsr membrane-targeting domain, was amplified from pDHL876 with primers DHL_P583_F and DHL_P584_R. The second part, which corresponds to ubiquitin, the *rpoS* coding region and FRT-flanked Kan resistance marker, was also amplified from pDHL876 using primers DHL_P585_F and DHL_P586_R. The third fragment corresponds to mEGFP and was amplified from pDHL962 with primers DHL_P587_F and DHL_P588_R.

pDHL1024 was built by isothermal assembly[181]. Three PCR fragments were ligated into SacI/XmaI-digested pUC19. The first PCR fragment corresponds to the first 90 amino acids of RpoS (i.e. $RpoS_{(1-90)}$) and was PCR amplified from pDHL20 with primers DHL_P466_F and DHL_P596_R. The second fragment corresponds to mEGFP and was amplified from pDHL962 with primers DHL_P597_F and DHL_P598_R. The third fragment corresponds to the last 240 amino acids of RpoS (i.e. $RpoS_{(91-330)}$) and was PCR amplified from pDHL20 with primers DHL_P599_F and DHL_P442_R.

pDHL1025 was built with isothermal assembly[181]. Two PCR fragments were ligated into pUC19, which was previously digested with SacI and XmaI. The first fragment corresponds to mEGFP and was amplified from pDHL962 with primers DHL_P600_F and DHL_P601_R. The second fragment contains the *rpoS* open-reading frame and was PCR amplified from pDHL20 with primers DHL_P602_F and DHL_P442_R.

pDHL1100 was built by amplifying the *sprE*-3xFLAG tag from genomic DNA of strain DHL331 (MC4100 *sprE*-3xFLAG) with primers DHL_P659_F and DHL_P670_R. The

PCR fragment was then purified and ligated with isothermal assembly[181] into pPM16, with was digested with BamHI/HindIII.

Primers

All primers were purchased from Integrated DNA Technologies, Inc. (IDT) and are listed in **Table 4.2**.

Table 4.2

Primer	Sequence
DHL_P13_F	cctt-gaattc-aacgtgaggaaatacctggatttttcc
DHL_P14_R	aagg-gagctc-cgtggtatcttccggaccgttc
DHL_P15_F	cctt-gagctc-agtaaaggagaagaacttttcactggagttg
DHL_P16_R	aagg-cccggg-ggcggatttgtcctactcaggag
DHL_P46_R	agaacagcctgccagccatagc
DHL_P47_F	ttgcgttttcccttgtccagatag
DHL_P55_F	cg-ggatcc-aacgtgaggaaatacctggatttttcc
DHL_P57_R	ccg-ctcgag-aaggtggctcctacccgtgatcc
DHL_P79_R	cggtgccctgaatgaactgc
DHL_P80_F	gcccagtcatagccgaatagcc
DHL_P81_R	gcgacgatagtaccaccagcc
DHL_P82_F	aaatccgtaaacccgctgcg
DHL_P83_R	tccgatgggcatcggacc
DHL_P88_F	tgtcataaagttgtcacggccg
DHL_P89_R	aacagcaaaaaaaccacccgg
DHL_P95_F	aagaagttattgaagcatcctcgtcagtaaaaagttaatcttttcaacagggatctatcaacagga gtccaagcg
DHL_P96_R	cagcaaaaaaaccacccggcagcgaaaattcactgccgggcgcggttttaaccagaacagcc cgtttgcg
DHL_P101_F	tcgaaagaactgtgtgcgcagg
DHL_P102_R	aggacatagcgttggctacccg
DHL_P103_F	ttacctgtccacacaatctgccc
DHL_P104_R	aaagttctctcggcagcgcc
DHL_P105_F	cctt-gagctc-gactacaaagaccatgacggtgattataaag
DHL_P106_R	aagg-cccggg-ccatatgaatatcctccttagttcctattcc
DHL_P107_F	gccaaatatggggaaccggtggtcgactgcgcttgatgttgtctgcagaa-gactacaaagaccatgacggtgattataaag
DHL_P108_R	agccgacattagcaggtaatgcaaatttagcccgcgttatcgtttgctca-ccatatgaatatcctccttagttcctattcc
DHL_P118_F	cg-ggatcc-ggatcaggtgcaacctttcacc
DHL_P119_R	ccg-ctcgag-gttctctcccctgactggctttactc
DHL_P120_F	ttaattatcgtcaattggttgccgc

Table 4.2 (Continued).

DHL_P121_R	ggggatcttgaagttcctattccg
DHL_P130_F	gcgaaatcctgcaaacgcaggggctgaatatcgaagcgctgttccgcgag-gactacaaagaccatgacggtgattataaag
DHL_P131_R	gaaaaggccagcctcgcttgagactggcctttctgacagatgcttactta-ccatatgaatatcctccttagttcctattcc
DHL_P134_F	aacggtccggaagataccacg
DHL_P135_R	gctcctggacgtagccttcgg
DHL_P144_F	gccaaatatggggaaccggtggtcgactgcgcttgatgttgtctgcagaa-ggtggtttatccagcaagggcg
DHL_P145_R	agccgacattagcaggtaatgcaaatttagcccgcgttatcgtttgctca-cgctatgacggaacaggtattcgc
DHL_P161_F	ggaattc-gagctc-gatcttttctacggggtctgacgc
DHL_P162_R	ggaattc-agatct-cttagacgtcaggtggcacttttcg
DHL_P163_F	ccg-ctcgag-agatcctctagatttaagaaggagatatacat-atggttagtaaaggagaagaaaataacatgg
DHL_P168_F	gccaaatatggggaaccggtggtcgactgcgcttgatgttgtctgcagaa-agcggtggcggtggcagtaa
DHL_P169_R	agccgacattagcaggtaatgcaaatttagcccgcgttatcgtttgctca-attccggggatccgtcgacc
DHL_P172_R	ggaagg-cctagg-tctagggcggcggatttgtcc
DHL_P271_F	cggcagacaacgaaaccgttatg
DHL_P272_R	tgtgccattgtgatgggctctg
DHL_P295_F	cctt-gagctc-agcggtggcggtggc-agtaaa-gtgagcaagggcgaggagctg
DHL_P404_F	ccg-ctcgag-agatcctctagatttaagaaggagatatacat-atggtgtctgagctgattaaggagaacatg
DHL_P405_R	ggaattc-cctgcagg-tctggacat-tta-tctgtgccccagtttgctaggg
DHL_P433_F	cgacgttgtaaaacgacggccagtgaattcatgttaaaacgtatcaaaattgtgaccagc
DHL_P434_R	aactccagtgaaaagttcttctcctttactggataaaccacctcttagccttagcacaag
DHL_P435_F	ttgtgctaaggctaagaggtggtttatccagtaaaggagaagaacttttcactggagttg
DHL_P436_R	acgaaaatctgcatgccaccgccaccgcttttgtatagttcatccatgccatgtgtaatc
DHL_P437_F	atggatgaactatacaaaagcggtggcggtggcatgcagattttcgtcaagactttgacc
DHL_P438_R	taaatcatgaactttcagcgtattctgactcataccaccgcggagccttagcacaagatg
DHL_P439_F	ccttacatcttgtgctaaggctccgcggtggtatgagtcagaatacgctgaaagttcatg
DHL_P442_R	gtataggaacttcgaagcagctccagcctacaccccgggttactcgcggaacagcgcttc
DHL_P456_F	gcgaaatcctgcaaacgcaggggctgaatatcgaagcgctgttccgcgag-agcggtggcggtggcagtaa
DHL_P457_R	acagaaaaggccagcctcgcttgagactggcctttctgacagatgcttac-attccggggatccgtcgacc
DHL_P466_F	gttgtaaaacgacggccagtgaattcgagctcatgagtcagaatacgctgaaagttcatg
DHL_P467_R	caactccagtgaaaagttcttctcctttactaccagagccggcgacatctccacgcagtg
DHL_P468_F	cactgcgtggagatgtcgccggctctggtagtaaaggagaagaacttttcactggagttg
DHL_P469_R	tcgatcatccggcggcgagagccggaacctttgtatagttcatccatgccatgtgtaatc
DHL_P470_F	gggattacacatggcatggatgaactatacaaaggttccggctctcgccgccggatgatc
DHL_P482_F	gttgtaaaacgacggccagtgaattcgagctcatgagtaaaggagaagaacttttcactg
DHL_P483_R	agcgtattctgactgccaccgccaccgcttttgtatagttcatccatgccatgtgtaatc
DHL_P484_F	gatgaactatacaaaagcggtggcggtggcagtcagaatacgctgaaagttcatgattta
DHL_P487_F	tcgagagatcctctagatttaagaaggagatatacatatggttagtaaag

Table 4.2 (Continued).

DHL_P488_R	cttctcctttactaaccatatgtatatctccttcttaaatctagaggatctc
DHL_P489_F	gagaagaaaataacatggcactgattaaggagaacatgcacatgaagct
DHL_P490_R	gtacagcttcatgtgcatgttctccttaatcagtgccatgttatttt
DHL_P540_F	cttttgcttgaatgttccgtcaagggatcacgggtaggagccaccttatg-attccggggatccgtcgacc
DHL_P541_R	tgagactggcctttctgacagatgcttacttactcgcggaacagcgcttc-tgtaggctggagctgcttcg
DHL_P543_R	aagg-cccggg-tta-cttgtacagctcgtccatgccgag
DHL_P583_F	ggtcgacggatccccggaat
DHL_P584_R	cccggtgaacagctcctcgcccttgctcac-ggataaaccacctcttagccttagcacaag
DHL_P585_F	cgccgggatcactctcggcatggacgagctgtacaag-agcggtggcggtggcatg
DHL_P586_R	gtatgttgtgtggaattgtgagcgg
DHL_P587_F	gttgatcatcttgtgctaaggctaagaggtggtttatcc-gtgagcaagggcgaggagctg
DHL_P588_R	aaagtcttgacgaaaatctgcatgccaccgccaccgct-cttgtacagctcgtccatgccg
DHL_P595_F	cttttgcttgaatgttccgtcaagggatcacgggtaggagccaccttatg-gtgagcaagggcgaggagctg
DHL_P596_R	gtgaacagctcctcgcccttgctcacaccagagccggcgacatctccacgcagtg
DHL_P597_F	tcgcgcactgcgtggagatgtcgccggctctggtgtgagcaagggcgaggagctg
DHL_P598_R	tctcgatcatccggcggcgagagccggaacccttgtacagctcgtccatgccgag
DHL_P599_F	cactctcggcatggacgagctgtacaagggttccggctctcgccgccggatgatc
DHL_P600_F	ttgtaaaacgacggccagtgaattcgagctcatggtgagcaagggcgaggagctg
DHL_P601_R	tcagcgtattctgactgccaccgccaccgctcttgtacagctcgtccatgccgag
DHL_P602_F	gctgtacaagagcggtggcggtggcagtcagaatacgctgaaagttcatgattta
DHL_P647_F	cttttgcttgaatgttccgtcaagggatcacgggtaggagccaccttatg-agtaaaggagaagaacttttcactggagttg
DHL_P648_F	aggcttttgcttgaatgttccgtcaagggatcacgggtaggagccaccttatgagtcagaatacgctgaaagttcatg
DHL_P659_F	cactgaccgaattcattaaagaggagaaaggatcc-atgacgcagccattggtcgg
DHL_P660_F	aatttgatgagaacggagttgaggtttttgacgaaaaggccttagtagaa-gtgagcaagggcgaggagctg
DHL_P670_R	gatgcctctagactcagctaattaagctta-tttatcgtcgtcatctttgtagtcgatatc
DHL_P671_F	tggcaatagcatgccactattgagtaaagccagtcaggggagagaacatg-attccggggatccgtcgacc
DHL_P672_R	atgcaaatttagcccgcgttatcgtttgctcattctgcagacaacatcaa-tgtaggctggagctgcttcg
NL-71-2	tgacaaaccgtcatcttcggctacttttttctctgtcacagaatgaaaattcgacggccagtgaattcgagc
NL-72	tttaaataagcgttgatattcagtcaattacaaacattaataacgaagagaattccggggatccgtcgac
NL73_R	ggattcatcgactgtggccg

E. coli strains

All *E. coli* strains used in this study are listed in the **Table 4.3**.

Table 4.3

Strain	Description	Antibiotic marker	Reference
MC4100	Wild-type (wt) *E. coli* strain	–	Lab collection
DH5α	Standard cloning strain	–	Lab collection
CNP77	MC4100 *rpoS*::Kan	Kan	Celeste
CNP227	MC4100 *rpoS750*-GFP	–	Celeste Peterson (unpublished)
DHL51	MC4100 *rpoS*::Tn10	Tet	This study
DHL193	MC4100 pKD46	Amp	[205]
DHL199	MC4100 *phoA*::$P_{rpoS\ long}$-GFPmut2-Kan	Kan	This study
DHL222	MC4100 *phoA*::P_{rpoS}-*rpoS750*-Venus-FRT Kan FRT	Kan	This study
DHL241	MC4100 *sprE*::Tn10	Tet	This study
DHL290	MC4100 *sprE*-3xFLAG-FRT Kan FRT (integration strain was TB10)	Kan	This study
DHL295	MC4100 sprE-3xFLAG	–	This study
DHL307	MC4100 *sprE*-3xFLAG-FRT Kan FRT (integration strain was DHL193)	Kan	This study
DHL309	DHL51 *sprE*-3xFLAG-FRT Kan FRT	Kan	This study
DHL319	MC4100 pDHL304	Kan	This study
DHL320	DHL51 pDHL304	Kan	This study
DHL331	MC4100 *sprE*-3xFLAG	–	This study
DHL333	DHL51 *sprE*-3xFLAG	Tet	This study
DHL335	MC4100 *rpoS*-3xFLAG-FRT Kan FRT	Kan	This study
DHL352	DHL193 *sprE*-Venus (Sunney Xie)-T1 terminator-FRT Kan FRT	Kan	This study
DHL363	DHL222 *sprE*::Tn10	Tet	This study
DHL394	DHL193 *sprE*-Venus-T1 terminator-FRT Kan FRT	Kan	This study
DHL399	MC4100 *sprE*-Venus-T1 terminator-FRT Kan FRT	Kan	This study
DHL438	DHL331 *rpoS*-3xFLAG-FRT Kan FRT	Kan	This study
DHL451	MC4100 pDHL450	Amp	This study
DHL452	DHL199 pDHL450	Amp	This study
DHL749	MC4100 ara+ pKD46	Amp	[205]
DHL812	MC4100 *sprE*-Venus-T1 terminator	–	This study
DHL813	DHL812 pDHL450		This study
DHL822	DHL199 pDHL819	Amp, Kan	This study
DHL870	MC4100 pDHL878	Amp	This study
DHL892	MC4100 Δ$rpoS_{(41\text{-}330)}$	–	This study

Table 4.3 (Continued).

DHL900	DHL892 pKD46	Amp	This study
DHL921	DHL333 pDH450	Amp	This study
DHL950	DHL193 Δ*rpoS*::FRT Kan FRT	Kan	This study
DHL1015	MC4100 ara+ Δ*rpoS*	–	This study
DHL1016	DHL1015 pKD46	Amp	This study
DHL1021	DHL1016 P_{rpoS}-mEGFP-Ub-*rpoS*-FRT Kan FRT	Kan	This study
DHL1038	DHL1016 P_{rpoS}-mEGFP-*rpoS*-FRT Kan FRT	Amp, Kan	This study
DHL1040	DHL1016 P_{rpoS}-*rpoS*$_N$-mEGFP-*rpoS*$_C$-FRT Kan FRT	Amp, Kan	This study
DHL1042	DHL1016 P_{rpoS}-*rpoS*$_N$-mGFPmut3-*rpoS*$_C$-FRT Kan FRT	Amp, Kan	This study
DHL1044	DHL1016 P_{rpoS}-mGFPmut3-Ub-*rpoS*-FRT Kan FRT	Amp, Kan	This study
DHL1050	DHL749 P_{rpoS}-mEGFP-Ub-*rpoS*	–	This study
DHL1051	DHL1050 pJT184	Cm	This study
DHL1057	DH749 *rpoS*-mEGFP-FRT Kan FRT	Amp, Kan	This study
DHL1059	DH749 *rpoS*-mGFPmut3-FRT Kan FRT	Amp, Kan	This study
DHL1074	MC4100 P_{rpoS}-mEGFP-*rpoS*	–	This study
DHL1075	MC4100 P_{rpoS}-*rpoS*$_{(1-90)}$-mEGFP-*rpoS*$_{(91-330)}$	–	This study
DHL1076	MC4100 P_{rpoS}-*rpoS*$_{(1-90)}$-mGFPmut3-*rpoS*$_{(91-330)}$	–	This study
DHL1081	MC4100 P_{rpoS}-mGFPmut3-Ub-*rpoS*	–	This study
DHL1082	DHL749 *sspB*-3xFLAG-FRT Kan FRT	Kan	This study
DHL1088	DHL749 *glmS*::P_{RNAI}-hybrid mKate2-T1 terminator-FRT Kan FRT::*pstS*	Kan	This study
DHL1085	DHL749 Δ*sprE*::FRT Kan FRT	Kan	This study
DHL1095	DHL900 P_{rpoS}-*rpoS*$_{(1-32)}$-mEGFP-Ub-*rpoS*-FRT Kan FRT	Kan	This study
DHL1102	MC4100 *sspB*-3xFLAG	–	This study
DHL1103	MC4100 *glmS*::P_{RNAI}-hybrid mKate2-T1 terminator::*pstS*	–	This study
DHL1116	DHL331 *sspB*-3xFLAG	–	This study
DHL1117	DHL438 *sspB*-3xFLAG	–	This study
DHL1121	MC4100 *rpoS*-mEGFP	–	This study
DHL1122	MC4100 *rpoS*-mGFPmut3	–	This study
DHL1128	MC4100 P_{rpoS}-*rpoS*$_{(1-32)}$-mEGFP-Ub-*rpoS*	–	This study
DHL1140	DHL1128 pJT184	Cm	This study
DHL1170	DHL1102 Δ*rpoS*	–	This study
DHL1171	DHL1102 Δ*sprE*	–	This study
DHL1172	DHL1116 Δ*rpoS*	–	This study
DHL1173	DHL1171 *rpoS*-3xFLAG	–	This study
DHL1178	MC4100 pDHL878	Amp	This study
DHL1179	DHL1015 pDHL878	Amp	This study
DHL1180	DHL1074 pDHL878	Amp	This study
DHL1181	DHL1075 pDHL878	Amp	This study
DHL1182	DHL1076 pDHL878	Amp	This study

Table 4.3 (Continued).

DHL1184	DHL1081 pDHL878	Amp	This study
DHL1185	DHL1121 pDHL878	Amp	This study
DHL1186	DHL1122 pDHL878	Amp	This study
DHL1187	DHL1128 pDHL878	Amp	This study
DHL1188	DHL1140 pDHL878	Amp, Cm	This study
DHL1189	DHL1116 pDHL878	Amp	This study
DHL1190	DHL1117 pDHL878	Amp	This study
DHL1191	DHL1172 pDHL878	Amp	This study
DHL1200	DHL1050 pDHL878	Amp	This study
DHL1201	DHL1051 pDHL878	Amp	This study
DHL1202	DHL1184 pJT184	Amp, Cm	This study
DHL1205	DHL1103 *att*Tn7::P_{bolA}-GFPmut2-T1 terminator	–	This study
DHL1206	DHL1103 *att*Tn7::P_{lacZ}-GFPmut2-T1 terminator	–	This study
DHL1207	DHL1103 *att*Tn7::P_{rpsU}-GFPmut2-T1 terminator	–	This study
DHL1208	DHL1103 *att*Tn7::P_{rssAB}-GFPmut2-T1 terminator	–	This study
DHL1209	DHL1103 *att*Tn7::no promoter-GFPmut2-T1 terminator	–	This study
DHL1211	DHL1103 *att*Tn7::P_{wrbA}-GFPmut2-T1 terminator	–	This study
DHL1212	DHL1103 *att*Tn7::P_{rpoD}-GFPmut2-T1 terminator	–	This study
DHL1213	DHL1103 *att*Tn7::P_{sprE_long}-GFPmut2-T1 terminator	–	This study
TB10	MG1655 *nadA*::Tn10 λcI857 Δ(*cro-bioA*)	–	[258]

Strain DHL51 was built by P1 transducing the *rpoS*::Tn10(tet) allele into MC4100. The P1 lysate was obtained from Celeste Peterson (Suffolk University). P1 transduction was performed according to a protocol from the Sauer lab (http://openwetware.org/wiki/Sauer:P1vir_phage_transduction).

Strain DHL199 was built by integrating the *rpoS* promoter fusion (P_{rpoS_long}-GFPmut2) into the *phoA* locus of the *E. coli* chromosome. The P_{rpoS_long}-GFPmut2-Kan integration cassette was PCR amplified from plasmid pDHL138 with primers DHL_P95_F and DHL_P96_R, gel purified, DpnI digested and integrated into strain DHL193 with using the λRed method[172]. A more detailed protocol was previously

described[205]. Integration into the *phoA* locus was confirmed with primers DHL_P88_F and DHL_P89_R. The upstream integration scar was PCR verified with primers DHL_P101_F and DHL_P102_R, whereas primers DHL_P103_F and DHL_P104_R were used for the downstream integration site.

Strain DHL222 built by integration of the RpoS750-Venus degradation reporter into the *phoA* locus of the *E. coli* chromosome using the λRed method (see above). The P_{rpoS_long}-*rpoS*750-Venus-T1 terminator-FRT Kan FRT integration cassette was PCR amplified from pDHL39 with primers DHL_P93_F and DHL_P94_R. The upstream integration site was PCR verified with primers DHL_P101_F and DHL_P46_R; and the downstream integration site was verified with primer set DHL_P47_F and DHL_P104_R.

Strain DHL241 was built by P1 transducing the *sprE*::Tn10(tet) allele into MC4100. The P1 lysate was obtained from Celeste Peterson (Suffolk University).

Strain DHL290 was built amplifying the 3xFLAG-FRT Kan FRT cassette from pDHL229 with primers DHL_P107_F and DHL_P108_R. The purified integration cassette was integrated into strain TB10. Tagging of *sprE* with the 3xFLAG tag was PCR verified with primers DHL_P120_F and DHL_P121_R.

Strain DHL295 is identical to DHL290 except that the FRT-flanked Kan marked was removed with pCP20.

Strain DHL307 was constructed identical to strain DHL290 except that the PCR product was integrated in strain DHL193 and not strain TB10.

Strain DHL309 corresponds to strain DHL51 with the sprE-3xFLAG allele, which was P1 transduced into this strain. The P1 lysate was made from strain DHL307.

Strain DHL331 is identical to strain DHL307 except that the FRT-flanked Kan marker was removed with pCP20.

Strain DHL333 is identical to strain DHL309 except that the FRT-flanked Kan marker was removed with pCP20.

Strain DHL335 was build by PCR amplifying the 3xFLAG tag-FRT Kan FRT integration cassette from plasmid pDHL229 with primer DHL_P130F and DHL_P131_R. The chromosomal integration was done in strain DHL193. Integration was verified by PCR with primers DHL_P134_F and DHL_P121_R for the upstream integration scar and primers DHL_P79_R and DHL_P83_R for the downstream integration scar.

Strain DHL352 was built by PCR amplifying the Venus-FRT Kan FRT integration cassette from plasmid pVS152Tsr[191] with primers DHL_P144_F and DHL_P145_R followed by integration into strain DHL193. The upstream integration scar was PCR verified with primers DHL_P120_F and DHL_P135_R, whereas the downstream integration scar was verified with primers DHL_P121_R and DHL_P81_R.

Strain DHL363 corresponds to strain DHL222 with the *sprE*::Tn10(tet) allele, which was transduced into DHL222 by P1. The P1 lysate of *sprE*::Tn10(tet) was made from strain DHL241.

Strain DHL394 was constructed by PCR amplifying the Venus-T1 terminator-FRT Kan FRT cassette from pDHL146 with primers DHL_P168_F and DHL_P169_R and integrated into strain DHL193. The upstream integration scar was PCR verified with primers DHL_P120_F and DHL_P79_R, whereas the downstream integration scar was verified with primers DHL_P80_F and DHL_P81_R.

Strain DHL399 is identical to strain DHL394 except that the sprE-Venus-T1 terminator-FRT Kan FRT cassette was P1 transduced from DHL193 (i.e. after the λRed integration) into fresh MC4100.

Strain DHL438 was build by transducing the *rpoS*-3xFLAG-FRT Kan FRT allele from strain DHL335 into strain DHL331.

Strain DHL812 is identical to strain DHL399 except that the FRT-flanked Kan marker was removed with pCP20.

Strain DHL892 was built by PCR amplifying the Δ*rpoS*::FRT Kan FRT deletion allele from the KEIO collection[259] using genomic DNA of strain JW5437-1 as the template and PCR primers DHL_P82_F and DHL_P83_R. The Δ*rpoS*::FRT Kan FRT cassette was then integrated into strain DHL193 with the λRed technique (see above). The Δ*rpoS*::FRT Kan FRT allele was then P1 transduced into fresh MC4100 and the FRT-flanked Kan marker was removing with pCP20. The Δ*rpoS* deletion allele was then sequenced (Genewiz) with primers DHL_P82_F and DHL_P83_R. Sequencing revealed that this strain still has the first 40 amino acids of RpoS and hence I named this deletion allele $\Delta rpoS_{(41\text{-}330)}$.

Strain DHL950 was built by PCR amplifying the FRT-flanked Kan resistance marker from plasmid pKD13[172] with primers DHL_P540_F and DHL_P541_R. The PCR fragment was then integrated in strain DHL193 to knockout *rpoS*. The deletion was verified with primers DHL_P82_F and DHL_P83_R.

Strain DHL1015 was constructed by P1 transducing the Δ*rpoS*::FRT Kan FRT allele from DHL950 into MC4100 ara+. The FRT-flanked Kan marker was removed with pCP20. The Δ*rpoS* deletion allele was then sequenced (Genewiz) with primers DHL_P82_F and DHL_P83_R. The Δ*rpoS* deletion corresponds to a 34 amino acid long peptide, which has the first amino acid of RpoS (i.e. the methionine) and the last 6 amino acids (i.e. EALFRE). The sequence in between corresponds to the FRT site scar. This approach was introduced by the KEIO collection[259].

Strain1021 was built by PCR amplifying the mEGFP-Ub-*rpoS*-FRT Kan FRT integration cassette from pDHL1014 with primers DHL_P595_F and DHL_P457_R. The PCR product was then transformed into strain DHL1016 and integrated in the chromosome with the λRed method (see above). Integration was confirmed by sequencing (Genewiz) the strain with primers DHL_P82_F and NL73_R.

Strain DHL1038 was built by PCR amplifying the mEGFP-*rpoS*-FRT Kan FRT integration cassette from pDHL1025 with primers DHL_P595_F and DHL_P457_R. The PCR product was then transformed into strain DHL1016 and integrated in the chromosome with the λRed method (see above). Integration was confirmed by sequencing (Genewiz) the strain with primers DHL_P82_F and NL73_R.

Strain DHL1040 was built by PCR amplifying the *rpoS*$_{(1-90)}$-mEGFP-*rpoS*$_{(91-330)}$-FRT Kan FRT integration cassette from pDHL1024 with DHL_P648_F and DHL_P457_R. The PCR product was then transformed into strain DHL1016 and integrated in the chromosome with the λRed method (see above). Integration was confirmed by sequencing (Genewiz) the strain with primers DHL_P82_F and NL73_R.

Strain DHL1042 was built by PCR amplifying the *rpoS*$_{(1-90)}$-mGFPmut3-*rpoS*$_{(91-330)}$-FRT Kan FRT integration cassette from pDHL896 with DHL_P648_F and DHL_P457_R. The PCR product was then transformed into strain DHL1016 and integrated in the chromosome with the λRed method (see above). Integration was confirmed by sequencing (Genewiz) the strain with primers DHL_P82_F and NL73_R.

Strain DHL1044 was built by PCR amplifying the mGFPmut3-Ub-*rpoS*-FRT Kan FRT integration cassette from pDHL876 with DHL_P647_F and DHL_P457_R. The PCR product was then transformed into strain DHL1016 and integrated in the chromosome with the λRed method (see above). Integration was confirmed by sequencing (Genewiz) the strain with primers DHL_P82_F and NL73_R.

Strain DHL1050 was built by PCR amplifying the mEGFP-Ub-*rpoS*-FRT Kan FRT integration cassette from pDHL1025 with primers DHL_P595_F and DHL_P457_R. The PCR product was then transformed into strain DHL1016 and integrated in the chromosome with the λRed method (see above).

Strain DHL1057 was built by PCR amplifying the mEGFP-FRT Kan FRT integration cassette from pDHL962 with primers DHL_P456_F and DHL_P457_R. The PCR product was transformed into strain DHL749 and integrated in the chromosome with the λRed method (see above). Integration was confirmed by sequencing (Genewiz) the strain with primers DHL_P82_F and NL73_R.

Strain DHL1059 was built by PCR amplifying the mGFPmut3-FRT Kan FRT integration cassette from pDHL580 with primers DHL_P456_F and DHL_P457_R. The PCR product was transformed into strain DHL749 and integrated in the chromosome with the λRed method (see above). Integration was confirmed by sequencing (Genewiz) the strain with primers DHL_P82_F and NL73_R.

Strain DHL1074 was constructed by P1 transducing the mEGFP-*rpoS*-FRT Kan FRT allele from strain DHL1038 into MC4100. The FRT-flanked Kan marker was removed with pCP20.

Strain DHL1075 was constructed by P1 transducing the *rpoS*$_{(1-90)}$-mEGFP-*rpoS*$_{(91-330)}$-FRT Kan FRT allele from strain DHL1040 into MC4100. The FRT-flanked Kan marker was removed with pCP20.

Strain DHL1076 was constructed by P1 transducing the *rpoS*$_{(1-90)}$-mGFPmut3-*rpoS*$_{(91-330)}$-FRT Kan FRT allele from strain DHL1042 into MC4100. The FRT-flanked Kan marker was removed with pCP20.

Strain DHL1081 was constructed by P1 transducing the mGFPmut3Ub-*rpoS*-FRT Kan FRT allele from strain DHL1044 into MC4100. The FRT-flanked Kan marker was removed with pCP20.

Strain DHL1082 was built by PCR amplifying the 3xFLAG-FRT Kan FRT integration cassette from pDHL229 with primers DHL_P657_F and DHL_P658_R. The PCR product was transformed into strain DHL749 and integrated in the chromosome with the λRed

method (see above). Integration was confirmed by sequencing (Genewiz) the strain with primers DHL_P271_F and DHL_P272_R.

Strain DHL1085 was built by PCR amplifying the FRT-flanked Kan resistance marker from plasmid pKD13[172] with DHL_P671_F and DHL_P672_R. The PCR fragment was then integrated in strain DHL749 to knockout *sprE*. The deletion was verified with primers DHL_P48_F and DHL_P81_R. The Δ*sprE*::FRT Kan FRT allele was then sequenced (Genewiz) with primers DHL_P48_F and DHL_P81_R. The Δ*sprE* deletion was constructed according to the KEIO collection[259] method: The *sprE* open-reading frame is replaced against a 34 amino acid long peptide, which has the first amino acid of SprE (i.e. the methionine) and the last 6 amino acids (i.e. MMLSAE).

Strain DHL1088 was built by PCR amplifying the P_{RNAI}-hybrid mKate2-FRT Kan FRT integration cassette from plasmid pNDL194 with primers NL-71-2 and NL-72. The PCR product was transformed into strain DHL749 and integrated in the chromosome with the λRed method (see above).

Strain DHL1095 was built by PCR amplifying the mEGFP-Ub-*rpoS*-FRT Kan FRT integration cassette from pDHL1014 with DHL_P660_F and DHL_P457_R. The PCR product was then transformed into strain DHL900 and integrated in the chromosome with the λRed method (see above). Integration was confirmed by sequencing (Genewiz) the strain with primers DHL_P82_F and NL73_R.

Strain DHL1102 was constructed by P1 transducing the *sspB*-3xFLAG-FRT Kan FRT allele from strain DHL1082 into MC4100. The FRT-flanked Kan marker was removed with pCP20.

Strain DHL1103 was constructed by P1 transducing the *glmS*::P_{RNAI}-hybrid mKate2-T1 terminator-FRT Kan FRT::*pstS* allele from strain DHL1088 into MC4100. The FRT-flanked Kan marker was removed with pCP20.

Strain DHL1116 was constructed by P1 transducing the *sspB*-3xFLAG-FRT Kan FRT allele from strain DHL1082 into DHL331. The FRT-flanked Kan marker was removed with pCP20.

Strain DHL1117 was constructed by P1 transducing the *sspB*-3xFLAG-FRT Kan FRT allele from strain DHL1082 into DHL438. The FRT-flanked Kan marker was removed with pCP20.

Strain DHL1121 was constructed by P1 transducing the *rpoS*-mEGFP-FRT Kan FRT allele from strain DHL1057 into MC4100. The FRT-flanked Kan marker was removed with pCP20.

Strain DHL1122 was constructed by P1 transducing the *rpoS*-mGFPmut3-FRT Kan FRT allele from strain DHL1059 into MC4100. The FRT-flanked Kan marker was removed with pCP20.

Strain DHL1128 was constructed by P1 transducing the P_{rpoS}-$rpoS_{(1-32)}$-mEGFP-Ub-*rpoS*-FRT Kan FRT allele from strain DHL1095 into MC4100. The FRT-flanked Kan marker was removed with pCP20.

Strain DHL1170 was constructed by P1 transducing the Δ*rpoS*::FRT Kan FRT allele from strain DHL950 into DHL1102. The FRT-flanked Kan marker was removed with pCP20.

Strain DHL1171 was constructed by P1 transducing the Δ*sprE*::FRT Kan FRT allele from strain DHL1085 into DHL1102. The FRT-flanked Kan marker was removed with pCP20.

Strain DHL1172 was constructed by P1 transducing the Δ*rpoS*::FRT Kan FRT allele from strain DHL950 into DHL1116. The FRT-flanked Kan marker was removed with pCP20.

Strain DHL1173 was constructed by P1 transducing the *rpoS*-3xFLAG-FRT Kan FRT allele from strain DHL335 into DHL1171. The FRT-flanked Kan marker was removed with pCP20.

Western blot against RpoS and SprE-3xFLAG

Western blot was performed as previously described[205]. Samples were normalized by OD_{600}. The polyclonal αGFP antibody (courtesy of Prof. P. Silver, Harvard Medical School) was used 1:1000 diluted. The monoclonal αRpoS antibody (Neoclone, W0009) was used 1:1000 diluted. The secondary antibodies were anti-mouse horseradish peroxidase (HRP)-linked secondary antibody (GE Healthcare, NA931) for RpoS and anti-rabbit HRP-linked secondary antibody (GE Healthcare, NA934) for the GFP and Venus fusions. The monoclonal αFLAG M2 antibody (Sigma-Aldrich, A8592) was used 1:5000 diluted and directly visualized with ECL reagent since the FLAG antibody is already HRP-coupled.

Epi-fluorescence microscopy of *E. coli* cells

Epi-fluorescence imaging was either performed with a Zeiss Axiovert 200M microscope or a Nikon Ti-E microscope. The Zeiss Axiovert 200M microscope was equipped with 100x phase objective (NA = 1.4, Zeiss), an automated x-y stage (ASI, MS-2000), a Lambda DG4 light source (Sutter) and an EM-CCD camera (Hamamatsu, C9100-02). The microscope was controlled with the software Axiovision (version 4.7).

The Nikon-Ti-E was equipped with a Perfect Focus System (PFS, Nikon), an Orca R2 (Hamamatsu) camera, a Scion CFW-1612M (Scion corporation) camera, a 100x Plan Apo objective (NA = 1.4, Nikon), an automated x−y stage (BioPrecision2 Inverted Stage, Ludl), an LED system (Spectra 7 light engine, Lumencor) and a white LED (pE-100, CoolLED) for phase-contrast imaging. The microscope was controlled by Micro-manager (http://www.micro-manager.org/) and custom-written Matlab scripts.

Evaluation of RpoS protein fusions by *sprE* promoter activity assay

Plate reader measurements were done with a Victor3 V 1420 plate reader (PerkinElmer) and appropriate filter sets for red fluorescence. Samples were grown to stationary phase in shaking 5-ml cultures. The overnight culture was transferred into a 96-well plate and OD600 and RFP intensities were measured. The strains were usually measured in 3 or 4 technical replicates, which all gave very similar values. Technical replicates were just averaged and not used for calculating error bars.

Mother machine experiment

The design and fabrication of the mother machine microfluidic chip will be described elsewhere (Nate Lord, manuscript in preparation).

V. Bibliography

1. Ferrell, J. E. & Machleder, E. M. The biochemical basis of an all-or-none cell fate switch in Xenopus oocytes. *Science* **280**, 895–898 (1998).

2. Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I. & van Oudenaarden, A. Multistability in the lactose utilization network of Escherichia coli. *Nature* **427**, 737–740 (2004).

3. Lahav, G. *et al.* Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nat Genet* **36**, 147–150 (2004).

4. Cai, L., Dalal, C. K. & Elowitz, M. B. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* **455**, 485–490 (2008).

5. Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. Bacterial persistence as a phenotypic switch. *Science* **305**, 1622–1625 (2004).

6. Süel, G. M., Garcia-Ojalvo, J., Liberman, L. M. & Elowitz, M. B. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* **440**, 545–550 (2006).

7. Weber, H., Polen, T., Heuveling, J., Wendisch, V. F. & Hengge-Aronis, R. Genome-wide analysis of the general stress response network in Escherichia coli: sigmaS-dependent genes, promoters, and sigma factor selectivity. *J Bacteriol* **187**, 1591–1603 (2005).

8. Hengge-Aronis, R. Signal Transduction and Regulatory Mechanisms Involved in Control of the σS (RpoS) Subunit of RNA …. *Microbiology and Molecular Biology Reviews* (2002).

9. Battesti, A., Majdalani, N. & Gottesman, S. The RpoS-mediated general stress response in Escherichia coli. *Annu. Rev. Microbiol.* (2011).

10. Pruteanu, M. & Hengge-Aronis, R. The cellular level of the recognition factor RssB is rate-limiting for sigmaS proteolysis: implications for RssB regulation and signal transduction in sigmaS turnover in Escherichia coli. *Mol Microbiol* **45**, 1701–1713 (2002).

11. Ruiz, N., Peterson, C. N. & Silhavy, T. J. RpoS-dependent transcriptional control of sprE: regulatory feedback loop. *J Bacteriol* **183**, 5974–5981 (2001).

12. Wu, X., Bayle, J. H., Olson, D. & Levine, A. J. The p53-mdm-2 autoregulatory feedback loop. *Genes Dev* **7**, 1126–1132 (1993).

13. Hoffmann, A., Levchenko, A., Scott, M. L. & Baltimore, D. The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. *Science* **298**, 1241–1245 (2002).

14. Yura, T., Nagai, H. & Mori, H. Regulation of the heat-shock response in bacteria. *Annu. Rev. Microbiol.* **47**, 321–350 (1993).

15. El-Samad, H. Regulated Degradation Is a Mechanism for Suppressing Stochastic Fluctuations in Gene Regulatory Networks. *Biophys J* **90**, 3749–3761 (2006).

16. Gibson, K. E. & Silhavy, T. J. SprE levels are growth phase regulated in a sigma(S)-dependent manner at the level of translation. *J Bacteriol* **182**, 4117–4120 (2000).

17. Lestas, I., Vinnicombe, G. & Paulsson, J. Fundamental limits on the suppression of molecular fluctuations. *Nature* **467**, 174–178 (2010).

18. Fontaine, F., Stewart, E. J., Lindner, A. B. & Taddei, F. Mutations in two global regulators lower individual mortality in Escherichia coli. *Mol Microbiol* **0**, 071202153222001–??? (2007).

19. Veening, J.-W., Smits, W. K. & Kuipers, O. P. Bistability, epigenetics, and bet-hedging in bacteria. *Annu. Rev. Microbiol.* **62**, 193–210 (2008).

20. Kitagawa, M. *et al.* Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research. *DNA Research* **12**, 291 (2006).

21. Taniguchi, Y. *et al.* Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010).

22. Winkler, J. *et al.* Quantitative and spatio-temporal features of protein aggregation in Escherichia coli and consequences on protein quality control and cellular ageing. *EMBO J* (2010).doi:10.1038/emboj.2009.412

23. Kain, J., He, G. G. & Losick, R. Polar localization and compartmentalization of ClpP proteases during growth and sporulation in Bacillus subtilis. *J Bacteriol* **190**, 6749–6757 (2008).

24. Kirstein, J., Strahl, H., Molière, N., Hamoen, L. W. & Turgay, K. Localization of general and regulatory proteolysis in Bacillus subtilis cells. *Mol Microbiol* **70**, 682–694 (2008).

25. Kirstein, J., Molière, N., Dougan, D. A. & Turgay, K. Adapting the machine: adaptor proteins for Hsp100/Clp and AAA+ proteases. *Nat Rev Micro* **7**, 589–599 (2009).

26. Kirstein, J. *et al.* The antibiotic ADEP reprogrammes ClpP, switching it from a regulated to an uncontrolled protease. *EMBO Molecular Medicine* **1**, 37–49 (2009).

27. Simmons, L. A., Grossman, A. D. & Walker, G. C. Clp and Lon proteases occupy distinct subcellular positions in Bacillus subtilis. *J Bacteriol* **190**, 6758–6768 (2008).

28. Werner, J. *et al.* Quantitative genome-scale analysis of protein localization in an asymmetric bacterium. *Proceedings of the National Academy of Sciences* **106**, 7858 (2009).

29. McGrath, P. T., Iniesta, A. A., Ryan, K. R., Shapiro, L. & McAdams, H. H. A dynamically localized protease complex and a polar specificity factor control a cell cycle master regulator. *Cell* **124**, 535–547 (2006).

30. Dziedzic, R. *et al.* Mycobacterium tuberculosis ClpX Interacts with FtsZ and Interferes with FtsZ Assembly. *PLoS ONE* **5**, e11058 (2010).

31. Zacharias, D. A., Violin, J. D., Newton, A. C. & Tsien, R. Y. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **296**, 913–916 (2002).

32. Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025–1036 (2005).

33. Zenklusen, D., Larson, D. R. & Singer, R. H. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol* **15**, 1263–1271 (2008).

34. Volkmer, B. & Heinemann, M. Condition-dependent cell volume and concentration of Escherichia coli to facilitate data conversion for systems biology modeling. *PLoS ONE* **6**, e23126 (2011).

35. Milo, R., Jorgensen, P., Moran, U., Weber, G. & Springer, M. BioNumbers--the database of key numbers in molecular and cell biology. *Nucleic Acids Res* **38**, D750–3 (2010).

36. Rudner, D. Z. & Losick, R. Protein subcellular localization in bacteria. *Cold Spring Harb Perspect Biol* **2**, a000307 (2010).

37. Shapiro, L., McAdams, H. H. & Losick, R. Why and how bacteria localize proteins. *Science* **326**, 1225–1228 (2009).

38. Bi, E. F. & Lutkenhaus, J. FtsZ ring structure associated with division in Escherichia coli. *Nature* **354**, 161–164 (1991).

39. Ma, X., Ehrhardt, D. W. & Margolin, W. Colocalization of cell division proteins FtsZ and FtsA to cytoskeletal structures in living Escherichia coli cells by using green fluorescent protein. *Proc Natl Acad Sci USA* **93**, 12998–13003 (1996).

40. Charbon, G. *et al.* Subcellular Protein Localization by Using a Genetically Encoded Fluorescent Amino Acid. *ChemBioChem* **12**, 1818–1821 (2011).

41. Fu, G. *et al.* In vivo structure of the E. coli FtsZ-ring revealed by photoactivated localization microscopy (PALM). *PLoS ONE* **5**, e12682 (2010).

42. Niu, L. & Yu, J. Investigating intracellular dynamics of FtsZ cytoskeleton with photoactivation single-molecule tracking. *Biophys J* **95**, 2009–2016 (2008).

43. Osawa, M., Anderson, D. E. & Erickson, H. P. Reconstitution of contractile FtsZ rings in liposomes. *Science* **320**, 792–794 (2008).

44. Raskin, D. M. & de Boer, P. A. Rapid pole-to-pole oscillation of a protein required for directing division to the middle of Escherichia coli. *Proc Natl Acad Sci USA* **96**, 4971–4976 (1999).

45. Hiraga, S., Ichinose, C., Niki, H. & Yamazoe, M. Cell cycle-dependent duplication and bidirectional migration of SeqA-associated DNA-protein complexes in E. coli. *Mol Cell* **1**, 381–387 (1998).

46. Chen, I. & Ting, A. Y. Site-specific labeling of proteins with small molecules in live cells. *Current Opinion in Biotechnology* **16**, 35–40 (2005).

47. Coons, A. H., Creech, H. J., Jones, R. N. & Berliner, E. The demonstration of pneumococcal antigen in tissues by the use of fluorescent antibody. *J. Immunol.* **45**, l942 (1942).

48. Fung, D. C. & Theriot, J. A. Imaging techniques in microbiology. *Current Opinion in Microbiology* **1**, 346–351 (1998).

49. Maddock, J. R. & Shapiro, L. Polar location of the chemoreceptor complex in the Escherichia coli cell. *Science* **259**, 1717–1723 (1993).

50. Schnell, U., Dijk, F., Sjollema, K. A. & Giepmans, B. N. G. Immunolabeling artifacts and the need for live-cell imaging. *Nat Meth* **9**, 152–158 (2012).

51. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. C. Green fluorescent protein as a marker for gene expression. *Science* **263**, 802–805 (1994).

52. Inouye, S. & Tsuji, F. I. Aequorea green fluorescent protein. Expression of the gene and fluorescence characteristics of the recombinant protein. *FEBS Lett* **341**, 277–280 (1994).

53. Wu, B., Piatkevich, K. D., Lionnet, T., Singer, R. H. & Verkhusha, V. V. Modern fluorescent proteins and imaging technologies to study gene expression, nuclear localization, and dynamics. *Current Opinion in Cell Biology* 1–8 (2011).doi:10.1016/j.ceb.2010.12.004

54. Lippincott-Schwartz, J., Snapp, E. & Kenworthy, A. Studying protein dynamics in living cells. *Nat Rev Mol Cell Biol* **2**, 444–456 (2001).

55. Hayashi, A. *et al.* Localization of gene products using a chromosomally tagged GFP-fusion library in the fission yeast Schizosaccharomyces pombe. *Genes Cells* **14**, 217–225 (2009).

56. Ding, D. Q. *et al.* Large-scale screening of intracellular protein localization in living fission yeast cells by the use of a GFP-fusion genomic DNA library. *Genes Cells* **5**, 169–190 (2000).

57. Matsuyama, A. *et al.* ORFeome cloning and global analysis of protein localization in the fission yeast Schizosaccharomyces pombe. *Nat Biotechnol* **24**, 841–847 (2006).

58. Huh, W.-K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).

59. Koroleva, O. A., Tomlinson, M. L., Leader, D., Shaw, P. & Doonan, J. H. High-throughput protein localization in Arabidopsis using Agrobacterium-mediated transient expression of GFP-ORF fusions. *Plant J.* **41**, 162–174 (2005).

60. Escobar, N. M. *et al.* High-throughput viral expression of cDNA-green fluorescent protein fusions reveals novel subcellular addresses and identifies unique proteins that interact with plasmodesmata. *Plant Cell* **15**, 1507–1523 (2003).

61. Morin, X., Daneman, R., Zavortink, M. & Chia, W. A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in Drosophila. *Proc Natl Acad Sci USA* **98**, 15050–15055 (2001).

62. Meissner, B. *et al.* Determining the sub-cellular localization of proteins within Caenorhabditis elegans body wall muscle. *PLoS ONE* **6**, e19937 (2011).

63. Simpson, J. C., Wellenreuther, R., Poustka, A., Pepperkok, R. & Wiemann, S. Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO Rep* **1**, 287–292 (2000).

64. Jarvik, J. W. *et al.* In vivo functional proteomics: mammalian genome annotation using CD-tagging. *BioTechniques* **33**, 852–4– 856– 858–60 passim (2002).

65. Frenkel-Morgenstern, M. *et al.* Dynamic Proteomics: a database for dynamics and localizations of endogenous fluorescently-tagged proteins in living human cells. *Nucleic Acids Res* **38**, D508–12 (2010).

66. Ormö, M. *et al.* Crystal structure of the Aequorea victoria green fluorescent protein. *Science* **273**, 1392–1395 (1996).

67. Yang, F., Moss, L. & Phillips, G., Jr The molecular structure of green fluorescent protein. *Nat Biotechnol* **14**, 1246–1251 (1996).

68. Shaner, N. C., Patterson, G. H. & Davidson, M. W. Advances in fluorescent protein technology. *Journal of Cell Science* **120**, 4247–4260 (2007).

69. Shaner, N. C., Steinbach, P. A. & Tsien, R. Y. A guide to choosing fluorescent proteins. *Nat Meth* **2**, 905–909 (2005).

70. Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol* **22**, 1567–1572 (2004).

71. Andresen, M., Schmitz-Salue, R. & Jakobs, S. Short tetracysteine tags to beta-tubulin demonstrate the significance of small labels for live cell imaging. *Mol Biol Cell* **15**, 5616–5622 (2004).

72. Snaith, H. A., Anders, A., Samejima, I. & Sawin, K. E. New and old reagents for fluorescent protein tagging of microtubules in fission yeast; experimental and critical evaluation. *Methods in Cell Biology* **97**, 147–172 (2010).

73. Baens, M. *et al.* The dark side of EGFP: defective polyubiquitination. *PLoS ONE* **1**, e54 (2006).

74. Yewdell, J. W., Lacsina, J. R., Rechsteiner, M. C. & Nicchitta, C. V. Out with the old, in with the new? Comparing methods for measuring protein degradation. *Cell Biol. Int.* **35**, 457–462 (2011).

75. Swulius, M. T. & Jensen, G. J. The helical MreB cytoskeleton in E. coli MC1000/pLE7 is an artifact of the N-terminal YFP tag. *J Bacteriol* (2012).doi:10.1128/JB.00505-12

76. Snapp, E. L. Fluorescent proteins: a cell biologist's user guide. *Trends Cell Biol* **19**, 649–655 (2009).

77. Werner, J. & Gitai, Z. High-Throughput Screening of Bacterial Protein Localization. *Meth Enzymol* **471**, 185–204 (2010).

78. Savage, D. F., Afonso, B., Chen, A. H. & Silver, P. A. Spatially ordered dynamics of the bacterial carbon fixation machinery. *Science* **327**, 1258–1261 (2010).

79. Haarer, B. K. & Pringle, J. R. Immunofluorescence localization of the Saccharomyces cerevisiae CDC12 gene product to the vicinity of the 10-nm filaments in the mother-bud neck. *Mol Cell Biol* **7**, 3678–3687 (1987).

80. McMurray, M. A. & Thorner, J. Septin stability and recycling during dynamic structural transitions in cell division and development. *Curr Biol* **18**, 1203–1208 (2008).

81. Katayama, H., Yamamoto, A., Mizushima, N., Yoshimori, T. & Miyawaki, A. GFP-like proteins stably accumulate in lysosomes. *Cell Struct Funct* 802040006 (2007).

82. Lauf, U., Lopez, P. & Falk, M. M. Expression of fluorescently tagged connexins: a novel approach to rescue function of oligomeric DsRed-tagged proteins. *FEBS Lett* **498**, 11–15 (2001).

83. Wang, W., Li, G.-W., Chen, C., Xie, X. S. & Zhuang, X. Chromosome organization by a nucleoid-associated protein in live bacteria. *Science* **333**, 1445–1449 (2011).

84. Yu, A. Y. H. & Houry, W. A. ClpP: a distinctive family of cylindrical energy-dependent serine proteases. *FEBS Lett* **581**, 3749–3757 (2007).

85. Kress, W., Maglica, Z. & Weber-Ban, E. Clp chaperone-proteases: structure and function. *Research in Microbiology* **160**, 618–628 (2009).

86. Striebel, F., Kress, W. & Weber-Ban, E. Controlled destruction: AAA+ ATPases in protein degradation from bacteria to eukaryotes. *Current Opinion in Structural Biology* **19**, 209–217 (2009).

87. Baker, T. A. & Sauer, R. T. ATP-dependent proteases of bacteria: recognition logic and operating principles. *Trends Biochem Sci* **31**, 647–653 (2006).

88. Sauer, R. T. *et al.* Sculpting the proteome with AAA(+) proteases and disassembly machines. *Cell* **119**, 9–18 (2004).

89. Gottesman, S. Proteolysis in bacterial regulatory circuits. *Annu Rev Cell Dev Biol* **19**, 565–587 (2003).

90. Keiler, K. C., Waller, P. R. & Sauer, R. T. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* **271**, 990–993 (1996).

91. Tu, G. F., Reid, G. E., Zhang, J. G., Moritz, R. L. & Simpson, R. J. C-terminal extension of truncated recombinant proteins in Escherichia coli with a 10Sa RNA decapeptide. *J Biol Chem* **270**, 9322–9326 (1995).

92. Gottesman, S., Roche, E., Zhou, Y. & Sauer, R. T. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev* **12**, 1338 (1998).

93. Farrell, C. M., Grossman, A. D. & Sauer, R. T. Cytoplasmic degradation of ssrA-tagged proteins. *Mol Microbiol* **57**, 1750–1761 (2005).

94. Lies, M. & Maurizi, M. R. Turnover of endogenous SsrA-tagged proteins mediated by ATP-dependent proteases in Escherichia coli. *Journal of Biological Chemistry* **283**, 22918 (2008).

95. Levchenko, I., Seidel, M., Sauer, R. T. & Baker, T. A. A specificity-enhancing factor for the ClpXP degradation machine. *Science* **289**, 2354–2356 (2000).

96. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).

97. Andersen, J. B. *et al.* New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl Environ Microbiol* **64**, 2240–2246 (1998).

98. Fung, E. *et al.* A synthetic gene-metabolic oscillator. *Nature* **435**, 118–122 (2005).

99. Grilly, C., Stricker, J., Pang, W. L., Bennett, M. R. & Hasty, J. A synthetic gene network for tuning protein degradation in Saccharomyces cerevisiae. *Mol Syst Biol* **3**, 127 (2007).

100. Chabot, J. R., Pedraza, J. M., Luitel, P. & van Oudenaarden, A. Stochastic gene expression out-of-steady-state in the cyanobacterial circadian clock. *Nature* **450**, 1249–1252 (2007).

101. Stricker, J. *et al.* A fast, robust and tunable synthetic gene oscillator. *Nature* **456**, 516–519 (2008).

102. Guet, C. C., Elowitz, M. B., Hsing, W. & Leibler, S. Combinatorial synthesis of genetic networks. *Science* **296**, 1466–1470 (2002).

103. Wong, W. W., Tsai, T. Y.-C. & Liao, J. C. Single-cell zeroth-order protein degradation enhances the robustness of synthetic oscillator. *Mol Syst Biol* **3**, 130 (2007).

104. Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H. & Weiss, R. A synthetic multicellular system for programmed pattern formation. *Nature* **434**, 1130–1134 (2005).

105. Kim, Y. I., Burton, R. E., Burton, B. M., Sauer, R. T. & Baker, T. A. Dynamics of substrate denaturation and translocation by the ClpXP degradation machine. *Mol Cell* **5**, 639–648 (2000).

106. Shin, Y. *et al.* Single-molecule denaturation and degradation of proteins by the AAA+ ClpXP protease. *Proc Natl Acad Sci USA* **106**, 19340–19345 (2009).

107. Austin, D. W. *et al.* Gene network shaping of inherent noise spectra. *Nature* **439**, 608–611 (2006).

108. Juillerat, A. *et al.* Directed evolution of O6-alkylguanine-DNA alkyltransferase for efficient labeling of fusion proteins with small molecules in vivo. *Chem Biol* **10**, 313–317 (2003).

109. Ohana, R. *et al.* HaloTag7: A genetically engineered tag that enhances bacterial expression of soluble proteins and improves protein purification. *Protein Expression and Purification* **68**, 110–120 (2009).

110. Los, G. V. & Wood, K. The HaloTag: a novel technology for cell imaging and protein analysis. *Methods Mol Biol* **356**, 195–208 (2007).

111. Los, G. V. *et al.* HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol* **3**, 373–382 (2008).

112. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* **24**, 79–88 (2006).

113. Ito, Y., Suzuki, M. & Husimi, Y. A novel mutant of green fluorescent protein with enhanced sensitivity for microanalysis at 488 nm excitation. *Biochemical and Biophysical Research Communications* **264**, 556–560 (1999).

114. Nagai, T. *et al.* A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat Biotechnol* **20**, 87–90 (2002).

115. Shcherbo, D. *et al.* Far-red fluorescent tags for protein imaging in living tissues. *Biochem J* **418**, 567–574 (2009).

116. Shaner, N. C. *et al.* Improving the photostability of bright monomeric orange and red fluorescent proteins. *Nat Meth* **5**, 545–551 (2008).

117. Elf, J., Li, G.-W. & Xie, X. S. Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell. *Science* **316**, 1191–1194 (2007).

118. Reyes-Lamothe, R., Sherratt, D. & Leake, M. Stoichiometry and architecture of active DNA replication machinery in Escherichia coli. *Science* **328**, 498 (2010).

119. Subach, O. M. *et al.* Conversion of red fluorescent protein into a bright blue probe. *Chem Biol* **15**, 1116–1124 (2008).

120. Ai, H.-W., Olenych, S. G., Wong, P., Davidson, M. W. & Campbell, R. E. Hue-shifted monomeric variants of Clavularia cyan fluorescent protein: identification of the molecular determinants of color and applications in fluorescence imaging. *BMC Biol.* **6**, 13 (2008).

121. Mckinney, S. A., Murphy, C. S., Hazelwood, K. L., Davidson, M. W. & Looger, L. L. A bright and photostable photoconvertible fluorescent protein. *Nat Meth* **6**, 131–133 (2009).

122. Chudakov, D. M., Lukyanov, S. & Lukyanov, K. A. Tracking intracellular protein movements using photoswitchable fluorescent proteins PS-CFP2 and Dendra2. *Nat Prot* **2**, 2024–2032 (2007).

123. Stiel, A. *et al.* 1.8 Å bright-state structure of the reversibly switchable fluorescent protein Dronpa guides the generation of fast switching variants. *Biochem J* **402**, 35 (2007).

124. Lawrence, M. S., Phillips, K. J. & Liu, D. R. Supercharging proteins can impart unusual resilience. *J Am Chem Soc* **129**, 10110–10112 (2007).

125. Costantini, L. M., Fossati, M., Francolini, M. & Snapp, E. L. Assessing the Tendency of Fluorescent Proteins to Oligomerize Under Physiologic Conditions. *Traffic (Copenhagen, Denmark)* (2012).doi:10.1111/j.1600-0854.2012.01336.x

126. Ulbrich, M. & Isacoff, E. Subunit counting in membrane-bound proteins. *Nat Meth* **4**, 319–321 (2007).

127. Ando, R., Mizuno, H. & Miyawaki, A. Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting. *Science* **306**, 1370 (2004).

128. Gurskaya, N. G. *et al.* Engineering of a monomeric green-to-red photoactivatable fluorescent protein induced by blue light. *Nat Biotechnol* **24**, 461–465 (2006).

129. Tsien, R. Y. The green fluorescent protein. *Annu Rev Biochem* **67**, 509–544 (1998).

130. Maurizi, M. R. *et al.* Sequence and structure of Clp P, the proteolytic component of the ATP-dependent Clp protease of Escherichia coli. *J Biol Chem* **265**, 12536–12545 (1990).

131. Tokunaga, M., Imamoto, N. & Sakata-Sogawa, K. Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat Meth* **5**, 159–161 (2008).

132. Crothers, D. & Metzger, H. The influence of polyvalency on the binding properties of antibodies. *Immunochemistry* **9**, 341–357 (1972).

133. Sauter, C., Basquin, J. & Suck, D. Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from Escherichia coli. *Nucleic Acids Res* **31**, 4091–4098 (2003).

134. Yoshimoto, T., Murayama, N., Honda, T., Tone, H. & Tsuru, D. Cloning and expression of aminopeptidase P gene from Escherichia coli HB101 and characterization of expressed enzyme. *Journal of Biochemistry* **104**, 93–97 (1988).

135. Wilce, M. C. *et al.* Structure and mechanism of a proline-specific aminopeptidase from Escherichia coli. *Proc Natl Acad Sci USA* **95**, 3472–3477 (1998).

136. Kitagawa, M., Miyakawa, M., Matsumura, Y. & Tsuchido, T. Escherichia coli small heat shock proteins, IbpA and IbpB, protect enzymes from inactivation by heat and oxidants. *Eur. J. Biochem.* **269**, 2907–2917 (2002).

137. Ratajczak, E. *et al.* IbpA the small heat shock protein from Escherichia coli forms fibrils in the absence of its cochaperone IbpB. *FEBS Lett* **584**, 2253–2257 (2010).

138. Buschmeier, B., Hengstenberg, W. & Deutscher, J. Purification and properties of 1-phosphofructokinase from Escherichia coli. *FEMS Microbiology Letters* **29**, 231–235 (1985).

139. Azam, T. A., Hiraga, S. & Ishihama, A. Two types of localization of the DNA-binding proteins within the Escherichia coli nucleoid. *Genes Cells* **5**, 613–626 (2000).

140. Diestra, E., Cayrol, B., Arluison, V. & Risco, C. Cellular electron microscopy imaging reveals the localization of the Hfq protein close to the bacterial membrane. *PLoS ONE* **4**, e8301 (2009).

141. Babic, A., Lindner, A. B., Vulic, M., Stewart, E. J. & Radman, M. Direct visualization of horizontal gene transfer. *Science* **319**, 1533–1536 (2008).

142. Lee, H., Kang, S., Bae, S. H., Choi, B. S. & Hwang, D. S. SeqA protein aggregation is necessary for SeqA function. *J Biol Chem* **276**, 34600–34606 (2001).

143. Lau, I. F. *et al.* Spatial and temporal organization of replicating Escherichia coli chromosomes. *Mol Microbiol* **49**, 731–743 (2003).

144. Golding, I. & Cox, E. C. RNA dynamics in live Escherichia coli cells. *Proc Natl Acad Sci USA* **101**, 11310–11315 (2004).

145. Wu, B., Chao, J. A. & Singer, R. H. Fluorescence Fluctuation Spectroscopy Enables Quantitative Imaging of Single mRNAs in Living Cells. *Biophys J* **102**, 2936–2944 (2012).

146. Chalfie, M. Green fluorescent protein. *Photochem Photobiol* **62**, 651–656 (1995).

147. Phillips, G. N. Structure and dynamics of green fluorescent protein. *Current Opinion in Structural Biology* **7**, 821–827 (1997).

148. Baird, G. S., Zacharias, D. A. & Tsien, R. Y. Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proc Natl Acad Sci USA* **97**, 11984–11989 (2000).

149. Jakobs, S., Subramaniam, V., Schönle, A., Jovin, T. M. & Hell, S. W. EGFP and DsRed expressing cultures of Escherichia coli imaged by confocal, two-photon and fluorescence lifetime microscopy. *FEBS Lett* **479**, 131–135 (2000).

150. Mizuno, H., Sawano, A., Eli, P., Hama, H. & Miyawaki, A. Red fluorescent protein from Discosoma as a fusion tag and a partner for fluorescence resonance energy transfer. *Biochemistry* **40**, 2502–2510 (2001).

151. Snapp, E. L. *et al.* Formation of stacked ER cisternae by low affinity protein interactions. *The Journal of Cell Biology* **163**, 257–269 (2003).

152. Söling, A., Simm, A. & Rainov, N. G. Intracellular localization of Herpes simplex virus type 1 thymidine kinase fused to different fluorescent proteins depends on choice of fluorescent tag. *FEBS Lett* **527**, 153–158 (2002).

153. Nakagawa, C., Nishimura, S., Senda-Murata, K. & Sugimoto, K. A Rapid and Simple Method of Evaluating the Dimeric Tendency of Fluorescent Proteins in Living Cells Using a Truncated Protein of Importin α as Fusion Tag. *Biosci Biotechnol Biochem* **76**, 388–390 (2012).

154. Wiedenmann, J. R., Oswald, F. & Nienhaus, G. U. Fluorescent proteins for live cell imaging: Opportunities, limitations, and challenges. *IUBMB Life* **61**, 1029–1042 (2009).

155. Day, R. N. & Davidson, M. W. The fluorescent protein palette: tools for cellular imaging. *Chemical Society reviews* **38**, 2887–2921 (2009).

156. Zacharias, D. A. Sticky caveats in an otherwise glowing report: oligomerizing fluorescent proteins and their use in cell biology. *Science's STKE* **2002**, pe23 (2002).

157. Miyawaki, A. Proteins on the move: insights gained from fluorescent protein technologies. *Nat Rev Mol Cell Biol* **12**, 656–668 (2011).

158. Yarbrough, D., Wachter, R. M., Kallio, K., Matz, M. V. & Remington, S. J. Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0-A resolution. *Proc Natl Acad Sci USA* **98**, 462–467 (2001).

159. Campbell, R. E. *et al.* A monomeric red fluorescent protein. *Proc Natl Acad Sci USA* **99**, 7877–7882 (2002).

160. Rekas, A., Alattia, J.-R., Nagai, T., Miyawaki, A. & Ikura, M. Crystal structure of venus, a yellow fluorescent protein with improved maturation and reduced environmental sensitivity. *J Biol Chem* **277**, 50573–50578 (2002).

161. Khmelinskii, A. *et al.* Tandem fluorescent protein timers for in vivo analysis of protein dynamics. *Nat Biotechnol* (2012).doi:10.1038/nbt.2281

162. Merzlyak, E. M. *et al.* Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nat Meth* **4**, 555–557 (2007).

163. Subach, F. V., Patterson, G. H., Renz, M., Lippincott-Schwartz, J. & Verkhusha, V. V. Bright monomeric photoactivatable red fluorescent protein for two-color super-resolution sptPALM of live cells. *J Am Chem Soc* **132**, 6481–6491 (2010).

164. Zhang, M. *et al.* Rational design of true monomeric and bright photoactivatable fluorescent proteins. *Nat Meth* (2012).doi:10.1038/nmeth.2021

165. Lowe, J. *et al.* Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 A resolution. *Science* **268**, 533–539 (1995).

166. Groll, M. *et al.* Structure of 20S proteasome from yeast at 2.4 A resolution. *Nature* 463–471 (1997).

167. Elowitz, M. B., Surette, M. G., Wolf, P. E., Stock, J. B. & Leibler, S. Protein mobility in the cytoplasm of Escherichia coli. *J Bacteriol* **181**, 197–203 (1999).

168. Swaminathan, R., Hoang, C. P. & Verkman, A. S. Photobleaching recovery and anisotropy decay of green fluorescent protein GFP-S65T in solution and cells: cytoplasmic viscosity probed by green fluorescent protein translational and rotational diffusion. *Biophys J* **72**, 1900–1907 (1997).

169. Erjavec, N., Larsson, L., Grantham, J. & Nyström, T. Accelerated aging and failure to segregate damaged proteins in Sir2 mutants can be suppressed by overproducing the protein aggregation-remodeling factor Hsp104p. *Genes Dev* **21**, 2410–2421 (2007).

170. Outeiro, T. F. & Lindquist, S. Yeast cells provide insight into alpha-synuclein biology and pathobiology. *Science* **302**, 1772–1775 (2003).

171. Huh, D. & Paulsson, J. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat Genet* (2010).

172. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci USA* **97**, 6640–6645 (2000).

173. McKenzie, G. J. & Craig, N. L. Fast, easy and efficient: site-specific insertion of transgenes into enterobacterial chromosomes using Tn7 without need for selection of the insertion event. *BMC Microbiol* **6**, 39–39 (2006).

174. Heikal, A., Hess, S., Baird, G., Tsien, R. & Webb, W. Molecular spectroscopy and dynamics of intrinsically fluorescent proteins: coral red (dsRed) and yellow (Citrine). *Proc Natl Acad Sci USA* **97**, 11996 (2000).

175. Griesbeck, O., Baird, G., Campbell, R., Zacharias, D. & Tsien, R. Reducing the environmental sensitivity of yellow fluorescent protein. *Journal of Biological Chemistry* **276**, 29188 (2001).

176. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* **25**, 1203–1210 (1997).

177. Lanzer, M. & Bujard, H. Promoters largely determine the efficiency of repressor action. *Proc Natl Acad Sci USA* **85**, 8973–8977 (1988).

178. Ajo-Franklin, C. M. *et al.* Rational design of memory in eukaryotic cells. *Genes Dev* **21**, 2271–2276 (2007).

179. Cormack, B. P., Valdivia, R. H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33–38 (1996).

180. Nguyen, A. W. & Daugherty, P. S. Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nat Biotechnol* **23**, 355–360 (2005).

181. Gibson, D. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Meth* **6**, 343–345 (2009).

182. Sun, X. *et al.* Development of SNAP-tag fluorogenic probes for wash-free fluorescence imaging. *Chembiochem : a European journal of chemical biology* **12**, 2217–2226 (2011).

183. Rosenfeld, N., Young, J. W., Alon, U., Surette, M. G. & Elowitz, M. B. Gene regulation at the single-cell level. *Science* **307**, 1962–1965 (2005).

184. Young, J. W. J. *et al.* Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nat Prot* **7**, 80–88 (2012).

185. Klein, T., Löschberger, A., Proppert, S. & Wolter, S. Live-cell dSTORM with SNAP-tag fusion proteins. *Nat Meth* (2011).

186. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Surette, M. G. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).

187. Rosenfeld, N., Perkins, T., Alon, U., Elowitz, M. B. & Swain, P. A fluctuation method to quantify in vivo fluorescence data. *Biophys J* **91**, 759–766 (2006).

188. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nat Genet* **31**, 69–73 (2002).

189. Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811–1814 (2004).

190. Cai, L., Friedman, N. & Xie, X. S. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358–362 (2006).

191. Yu, J., Xiao, J., Ren, X., Lao, K. & Xie, X. S. Probing gene expression in live cells, one protein molecule at a time. *Science* **311**, 1600–1603 (2006).

192. Bar-Even, A. *et al.* Noise in protein expression scales with natural protein abundance. *Nat Genet* **38**, 636–643 (2006).

193. Sigal, A. *et al.* Variability and memory of protein levels in human cells. *Nature* **444**, 643–646 (2006).

194. Newman, J. R. S. *et al.* Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441**, 840–846 (2006).

195. Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. Visualization of single RNA transcripts in situ. *Science* **280**, 585–590 (1998).

196. Raj, A., Van Den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Meth* **5**, 877–879 (2008).

197. Raj, A. & Tyagi, S. Detection of individual endogenous RNA transcripts in situ using multiple singly labeled probes. *Meth Enzymol* **472**, 365–386 (2010).

198. Zong, C., So, L., Sepúlveda, L., Skinner, S. & Golding, I. Lysogen stability is determined by the frequency of activity bursts from the fate-determining gene. *Mol Syst Biol* **6**, (2010).

199. So, L.-H. *et al.* General properties of transcriptional time series in Escherichia coli. *Nat Genet* **43**, 554–560 (2011).

200. Raj, A., Rifkin, S. A., Andersen, E. & van Oudenaarden, A. Variability in gene expression underlies incomplete penetrance. *Nature* **463**, 913–918 (2010).

201. Lubeck, E. & Cai, L. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nat Meth* (2012).doi:10.1038/nmeth.2069

202. Tal, S. & Paulsson, J. Evaluating quantitative methods for measuring plasmid copy numbers in single cells. *Plasmid* **67**, 167–173 (2012).

203. Bertrand, E. *et al.* Localization of ASH1 mRNA particles in living yeast. *Mol Cell* **2**, 437–445 (1998).

204. Peabody, D. S. & Ely, K. R. Control of translational repression by protein-protein interactions. *Nucleic Acids Res* **20**, 1649–1655 (1992).

205. Landgraf, D., Okumus, B., Chien, P., Baker, T. A. & Paulsson, J. Segregation of molecules at cell division reveals native protein localization. *Nat Meth* **9**, 480–482 (2012).

206. Guet, C. C. *et al.* Minimally invasive determination of mRNA concentration in single living bacteria. *Nucleic Acids Res* **36**, e73 (2008).

207. Le, T. T. *et al.* Real-time RNA profiling within a single bacterium. *Proc Natl Acad Sci USA* **102**, 9160–9164 (2005).

208. Chao, J. A., Patskovsky, Y., Almo, S. C. & Singer, R. H. Structural basis for the coevolution of a viral RNA-protein complex. *Nat Struct Mol Biol* **15**, 103–105 (2008).

209. Larson, D. R., Zenklusen, D., Wu, B., Chao, J. A. & Singer, R. H. Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science* **332**, 475–478 (2011).

210. Paige, J. S., Wu, K. Y. & Jaffrey, S. R. RNA Mimics of Green Fluorescent Protein. *Science* **333**, 642–646 (2011).

211. Jungmann, R. *et al.* Single-molecule kinetics and super-resolution microscopy by fluorescence imaging of transient binding on DNA origami. *Nano Lett.* **10**, 4756–4761 (2010).

212. Zenklusen, D. & Singer, R. H. Analyzing mRNA Expression Using Single mRNA Resolution Fluorescent In Situ Hybridization. *Meth Enzymol* (2010).

213. Llopis, P. *et al.* Spatial organization of the flow of genetic information in bacteria. *Nature* **466**, 77–81 (2010).

214. Shen, P. & Huang, H. V. Homologous recombination in Escherichia coli: dependence on substrate length and homology. *Genetics* **112**, 441–457 (1986).

215. Unger, M., Kartalov, E., Chiu, C. S., Lester, H. A. & Quake, S. R. Single-molecule fluorescence observed with mercury lamp illumination. *BioTechniques* **27**, 1008–1014 (1999).

216. Chen, I., Howarth, M., Lin, W. & Ting, A. Y. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat Meth* **2**, 99–104 (2005).

217. Delebecque, C. J., Lindner, A. B., Silver, P. A. & Aldaye, F. A. Organization of intracellular reactions with rationally designed RNA assemblies. *Science* **333**, 470–474 (2011).

218. Paulsson, J. Summing up the noise in gene networks. *Nature* **427**, 415–418 (2004).

219. Trcek, T. *et al.* Single-mRNA counting using fluorescent in situ hybridization in budding yeast. *Nat Prot* **7**, 408–419 (2012).

220. Cannon, W. B. Organization for physiological homeostasis. *Physiological Reviews* **9**, 399–431 (1929).

221. Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590–593 (2000).

222. Dublanche, Y., Michalodimitrakis, K., Kümmerer, N., Foglierini, M. & Serrano, L. Noise in transcription negative feedback loops: simulation and experimental analysis. *Mol Syst Biol* **2**, 41 (2006).

223. Løbner-Olesen, A. Distribution of minichromosomes in individual Escherichia coli cells: implications for replication control. *EMBO J* **18**, 1712–1721 (1999).

224. Paulsson, J. & Ehrenberg, M. Noise in a minimal regulatory network: plasmid copy number control. *Q Rev Biophys* **34**, 1–59 (2001).

225. Dong, T. & Schellhorn, H. E. Role of RpoS in virulence of pathogens. *Infect. Immun.* **78**, 887–897 (2010).

226. Klauck, E., Typas, A. & Hengge-Aronis, R. The sigmaS subunit of RNA polymerase as a signal integrator and network master regulator in the general stress response in Escherichia coli. *Science progress* **90**, 103–127 (2007).

227. Fredriksson, A. *et al.* Decline in ribosomal fidelity contributes to the accumulation and stabilization of the master stress response regulator sigmaS upon carbon starvation. *Genes Dev* **21**, 862–874 (2007).

228. Lange, R. & Hengge-Aronis, R. Identification of a central regulator of stationary-phase gene expression in Escherichia coli. *Mol Microbiol* **5**, 49–59 (1991).

229. Pratt, L. A. & Silhavy, T. J. The response regulator SprE controls the stability of RpoS. *Proc Natl Acad Sci USA* **93**, 2488–2492 (1996).

230. Muffler, A., Fischer, D., Altuvia, S., Storz, G. & Hengge-Aronis, R. The response regulator RssB controls stability of the sigma(S) subunit of RNA polymerase in Escherichia coli. *EMBO J* **15**, 1333–1339 (1996).

231. Nachman, I., Regev, A. & Ramanathan, S. Dissecting timing variability in yeast meiosis. *Cell* **131**, 544–556 (2007).

232. Alcedo, J., Zou, Y. & Noll, M. Posttranscriptional regulation of smoothened is part of a self-correcting mechanism in the Hedgehog signaling system. *Mol Cell* **6**, 457–465 (2000).

233. Lange, R. & Hengge-Aronis, R. The cellular concentration of the sigma S subunit of RNA polymerase in Escherichia coli is controlled at the levels of transcription, translation, and protein stability. *Genes Dev* **8**, 1600–1612 (1994).

234. Jishage, M. & Ishihama, A. Regulation of RNA polymerase sigma subunit synthesis in Escherichia coli: intracellular levels of sigma 70 and sigma 38. *J Bacteriol* **177**, 6832–6835 (1995).

235. Zhou, Y. & Gottesman, S. Regulation of proteolysis of the stationary-phase sigma factor RpoS. *J Bacteriol* **180**, 1154–1158 (1998).

236. Zgurskaya, H. I., Keyhan, M. & Matin, A. The sigma S level in starving Escherichia coli cells increases solely as a result of its increased stability, despite decreased synthesis. *Mol Microbiol* **24**, 643–651 (1997).

237. Stüdemann, A. *et al.* Sequential recognition of two distinct sites in sigma(S) by the proteolytic targeting factor RssB and ClpX. *EMBO J* **22**, 4111–4120 (2003).

238. Klauck, E., Lingnau, M. & Hengge-Aronis, R. Role of the response regulator RssB in sigma recognition and initiation of sigma proteolysis in Escherichia coli. *Mol Microbiol* **40**, 1381–1390 (2001).

239. Schweder, T., Lee, K. H., Lomovskaya, O. & Matin, A. Regulation of Escherichia coli starvation sigma factor (sigma s) by ClpXP protease. *J Bacteriol* **178**, 470–476 (1996).

240. Zhou, Y., Gottesman, S., Hoskins, J. R., Maurizi, M. R. & Wickner, S. The RssB response regulator directly targets sigma(S) for degradation by ClpXP. *Genes Dev* **15**, 627–637 (2001).

241. Becker, G., Klauck, E. & Hengge-Aronis, R. The response regulator RssB, a recognition factor for sigmaS proteolysis in Escherichia coli, can act like an anti-sigmaS factor. *Mol Microbiol* **35**, 657–666 (2000).

242. Peterson, C. N., Levchenko, I., Rabinowitz, J. D., Baker, T. A. & Silhavy, T. J. RpoS proteolysis is controlled directly by ATP levels in Escherichia coli. *Genes Dev* **26**, 548–553 (2012).

243. Paulsson, J. Models of stochastic gene expression. *Physics of Life Reviews* **2**, 157–175 (2005).

244. Bougdour, A., Wickner, S. & Gottesman, S. Modulating RssB activity: IraP, a novel regulator of sigma(S) stability in Escherichia coli. *Genes Dev* **20**, 884–897 (2006).

245. Xie, X. S., Choi, P. J., Li, G.-W., Lee, N. K. & Lia, G. Single-Molecule Approach to Molecular Biology in Living Bacterial Cells. *Annu. Rev. Biophys.* **37**, 417–444 (2008).

246. Muffler, A., Traulsen, D. D., Lange, R. & Hengge-Aronis, R. Posttranscriptional osmotic regulation of the sigma(s) subunit of RNA polymerase in Escherichia coli. *J Bacteriol* **178**, 1607–1613 (1996).

247. Sledjeski, D. D., Gupta, A. & Gottesman, S. The small RNA, DsrA, is essential for the low temperature expression of RpoS during exponential growth in Escherichia coli. *EMBO J* **15**, 3993–4000 (1996).

248. Ranquet, C. & Gottesman, S. Translational regulation of the Escherichia coli stress factor RpoS: a role for SsrA and Lon. *J Bacteriol* **189**, 4872–4879 (2007).

249. Jacobson, R. H., Zhang, X. J., DuBose, R. F. & Matthews, B. W. Three-dimensional structure of beta-galactosidase from E. coli. *Nature* **369**, 761–766 (1994).

250. Hensel, Z. *et al.* Stochastic expression dynamics of a transcription factor revealed by single-molecule noise analysis. *Nat Struct Mol Biol* **19**, 797–802 (2012).

251. Tobias, J. W., Shrader, T. E., Rocap, G. & Varshavsky, A. The N-end rule in bacteria. *Science* **254**, 1374–1377 (1991).

252. Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-sequence determinants of gene expression in Escherichia coli. *Science* **324**, 255–258 (2009).

253. Butt, T. R. *et al.* Ubiquitin fusion augments the yield of cloned gene products in Escherichia coli. *Proc Natl Acad Sci USA* **86**, 2540–2544 (1989).

254. Wang, P. *et al.* Robust growth of Escherichia coli. *Curr Biol* **20**, 1099–1103 (2010).

255. Megerle, J. A., Fritz, G., Gerland, U., Jung, K. & Rädler, J. O. Timing and dynamics of single cell gene expression in the arabinose utilization system. *Biophys J* **95**, 2103–2115 (2008).

256. Zaslaver, A. *et al.* A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. *Nat Meth* **3**, 623–628 (2006).

257. Uzzau, S., Figueroa-Bossi, N., Rubino, S. & Bossi, L. Epitope tagging of chromosomal genes in Salmonella. *Proc Natl Acad Sci USA* **98**, 15264–15269 (2001).

258. Johnson, J. E., Lackner, L. L., Hale, C. A. & de Boer, P. A. J. ZipA is required for targeting of DMinC/DicB, but not DMinC/MinD, complexes to septal ring assemblies in Escherichia coli. *J Bacteriol* **186**, 2418–2429 (2004).

259. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006.0008 (2006).